Natural Axiom System of Probability Theory
—— mathematical model of random universe

概率论自然公理系统

随 机 世 界 的 数 学 模 型

熊大国
著





清华大学出版社
http://www.tup.tsinghua.edu.cn

概率论自然公理系统

——随机世界的数学模型

熊大国　著

清 华 大 学 出 版 社

(京)新登字158号

内 容 提 要

本著作发展概率论的柯尔莫哥洛夫公理系统，建立现实随机世界的数学模型，形成自然公理系统。

全书共3章，其中第1章探讨事件和概率的本质以及它们在科学中的巨大作用；第2章介绍自然公理系统，给出6组公理，这些公理将随机世界抽象成因果空间，并在因果空间中引进两类"图形"——随机试验和概率空间，进一步引入条件概率和独立性等概念；第3章介绍随机变量的基本概念，论证它与所建立的公理系统之间的关系。

本书可供从事概率论及相关领域研究工作的人员参考。

书　名：概率论自然公理系统——随机世界的数学模型
作　者：熊大国
出版者：清华大学出版社（北京清华大学学研大厦，邮编100084）
http://www.tup.tsinghua.edu.cn
印刷者：清华大学印刷厂
发行者：新华书店总店北京发行所
开　本：850×1168　1/32　印张：6　字数：154千字
版　次：2000年11月第1版　2000年11月第1次印刷
书　号：ISBN 7-302-00802-7/O·114
印　数：0001～2000
定　价：14.00元

序

概率论是研究现实的随机世界，即自然界和人类社会中随机现象的一门数学分支。它的基础是由柯尔莫哥洛夫(Kolmogorov)公理系统(简称为K公理系统)奠定的。在K公理系统中概率空间$(\Omega,\mathscr{F},P)$是一切研究的出发点。其中Ω是某个集合，它的元素称为基本事件；$\mathscr{F}$是Ω上的σ域，它的元素称为事件；P是$\mathscr{F}$上的概率测度，数值$P(A)$是事件A出现可能性的度量。于是，集合论和测度论方法被硬性地搬到概率论中，并用它推演出概率论的理论体系。

这种研究模式至少存在五个缺点。① 它没有建立随机世界的整体形象。作为类比，初等几何建立了自己的整体形象——欧氏空间，使得图形及其度量的研究既形象又生动。试问：概率论为什么不这样做呢？② 由于概率空间被绝对化、从而切断了概率空间内外事件之间的联系，破坏了随机现象的直观性和形象性。③ 无法解释集合论方法为什么能用于研究随机现象。④ 没有讨论条件与概率之间的内在联系，很难揭示条件概率与独立性的本质。事实上，K公理系统引进的条件概率是混乱的。⑤ 概率论的核心内容被模糊。作为类比，不管现代几何学的发展如何多种多样，直观的欧氏几何总以某种形式存在于其他类型的几何之中，并且是产生这些几何的源泉。试问：概率论中的“欧氏几何”存在吗？是什么？

本书提出的概率论自然公理系统克服了上述缺点，发展出一种新的研究模式。自然公理系统把随机世界抽象成因果空间。因果空间中的事件不仅保持随机现象的直观性和形象性，而且获得多种表达形式。概率论中集合论方法是从多种表达式的转换中产生出来的。

从哲学的角度看，事件存在多种表达式是因果律中“因”和“果”在数学中的反映；关于因果律思维方式的数学抽象就是概率论中集合论方法和测度论方法。

因果空间中存在各种各样的“图形”。自然公理系统主要研究两类“图形”——随机试验和概率空间，其中的核心分别是离散型试验和波莱尔(Borel)试验，离散型概率空间和柯尔莫哥洛夫概率空间。在因果空间中不仅可以研究单个的“图形”，而且可以研究不同“图形”之间的关系和进行数学演算。

应用随机试验这个概念，我们揭示出概率、条件概率和独立性的本质。并且发现用拉东—尼科迪姆定理定义的条件概率实际上不是条件概率，而是取了一个不恰当的名称。

现实中的随机现象经过数学处理后几乎都可以用随机变量表达，并且应用柯尔莫哥洛夫概率空间的统计知识可以得到随机变量蕴含的全部统计知识和统计规律。于是，因果空间、离散型试验、波莱尔试验、离散型概率空间、柯尔莫哥洛夫概率空间和随机变量组成了概率论的核心内容。这些核心内容使人们认识到随机世界不是一个杂乱无章、难于捉摸的世界，而是一个条理清楚、按统计规律在不断变化的世界。

由于作者的知识和能力所限，本书又是首次阐明新思想的著作，因此书中必定存在不完善、疏漏、不妥和叙述含糊的地方，恳请读者指正和补充。

书稿完成后得到清华大学出版社的大力支持，使本书能够很快地献给读者。北京科技大学秦明达教授仔细地审阅了全稿，提出许多中肯的、宝贵的意见，使书稿质量大有提高。在写作过程中得到教育科学出版社郑桂泉副编审的许多帮助，加速了书稿的完成。在此，向他们表达最诚挚的谢意。

作　者

1999.11.6

目　录

绪　论

概率论的研究对象是现实世界中的随机现象(本书称之为随机世界)。柯尔莫哥洛夫(A. H. Kolmogorov)于1933年发表专著[1],建立了概率论的柯尔莫哥洛夫公理系统(简称为K公理系统)。K公理系统为现代概率论的蓬勃发展奠定牢固的基础,成为概率论发展史上的一个里程碑。

对K公理系统的批评主要来自贝叶斯(T. Bayes)学派。针对K公理系统的缺点和不足,菲纳特(B. de Finetti)提出7个讨论题[2]。一些作者建立与K公理系统稍有差别的公理系统[3~7]。综述性文献[8]介绍了建立主观概率公理系统的情况。

我们从另一方面加入K公理系统的批评者行列,指出如下7点不足或可商榷之处。

(1) 概率论公理系统应当完成一个宏伟的目标——建立随机世界的数学模型。

它应当把貌似杂乱无章的随机世界组织成一个既形象又严谨的逻辑体系。即是说,首先找出随机世界中几个最本质的属性,把它们抽象成公理系统,然后对公理系统进行逻辑推理和数学演算,最终推演出概率论的理论体系。目的是发现随机世界中更多更深刻的性质和规律。

但是,K公理系统没有提出这项任务。它只是研究随机世界的某个局部,并把它抽象成概率空间。概率空间之间的联系是不清楚的。

(2) 概率论为什么没有自身独特的研究场所?

随机现象在现实中如此具体、如此广泛。可是,研究它们的概

率论却没有为自己建筑“住房”——独特的研究场所[①]。这是一种不正常的现象。导致一些数学家认为:概率论成为测度论的一部分。究其原因,来自于在K公理系统中基本事件空间是一个任意的集合,概率空间是在这个集合上形成的一类特殊的测度空间。

(3) K公理系统破坏了随机世界的直观形象:

随机世界	K公理系统
① 必然事件皆相等,即必然事件是惟一的。 ② 任意两个事件都可以进行并、交、差运算。	① 不对。不同的概率空间有不同的必然事件。 ② 不对。只有同一概率空间中的两个事件才能进行并、交、差运算。

由此不难看出,K公理系统人为地切断了概率空间内外随机现象之间的联系,因而缩小了概率论和数理统计的研究范围。

(4) K公理系统中的事件(例如“明天晴”)为什么是基本事件的集合?而由基本事件形成的集合为什么有的是事件?有的又不是?

这个问题相当于菲纳特的第1个讨论题。进而我们可以用更哲理的方式询问以下问题。

(5) 集合论方法为什么会成为研究随机现象的基本方法?

(6) K公理系统中基本事件可以是事件,也可以不是事件。如果不是,试问它是什么?

(7) K公理系统没有探讨条件与概率之间的内在联系。条件概率和独立性的直观背景被掩盖,因而很难把它们推广到一般场合。

这个问题相当于菲纳特的第6个讨论题。

本书完满地解答了上述疑问,把K公理系统发展成概率论的

---

① 自然公理系统提供的“住房”是因果空间。

自然公理系统。自然公理系统把随机世界描述为一个像欧氏空间那样既形象又能进行数学演绎推理的空间——因果空间。表 0.1 用类比的方式显示自然公理系统的结构和意义。

表 0.1 概率论与欧氏几何学[9]类比表

学　科	欧氏几何学	概率论
直观的研究对象	现实世界中物体的形状	现实世界中的随机现象（随机世界）
公理化方法 元素 元素之间的关系	希尔伯特公理系统 点、直线、平面 5 组公理	自然公理系统 随机事件 6 组公理
公理化后的产物 研究对象	欧几里得空间 图形 图形的度量	因果空间 随机试验 概率空间
研究任务	欧氏几何学是研究图形和对图形进行度量的一门学科。它包括： ① 单个图形或单个度量化后图形的性质； ② 不同图形或不同度量化后图形之间的关系。	概率论是研究因果空间中随机试验和对试验进行赋概的一门学科。它包括： ① 单个随机试验或单个概率空间的性质； ② 不同随机试验或不同概率空间之间的关系。

自然公理系统具有两大特点。第 1 个特点是遵守现代化公理系统的要求，是一个形式化的公理系统。

形式化的含义是，首先找到概率论的基本对象——随机事件，把它作为不能定义的元素，并用符号 $A,B,C\cdots$ 表示它们；其次，把元素之间的关系抽象成 6 组公理，并且公理也被符号化；最后，以 6 组公理为基础，对符号进行逻辑推理和数学演算，得到新的符号组合——各种各样的结论，从而形成概率论的理论体系。由于在推理和演算时不必顾虑这些符号的具体意义，因而保证了结论的可

靠性。

第 2 个特点是具备古典公理系统所要求的直观性和形象性。在选择公理时力求简单、直观，的确不需证明就被人们所接受。

概率论有非常具体的研究对象——随机世界。因此，直观形象是概率论产生的源泉，存在的基础。它不仅启迪研究者的思维，帮助思考，而且不断地为概率论和数理统计提供新的研究任务、新的研究方向，并指明发展方向，使概率论充满了生机和活力，生长出许多边缘学科，形成非常庞大的数学分支。

在内容安排上作两点说明：

(1) 我们在引进某组公理之前先扼要地介绍这组公理的直观背景[①]。

虽然我们生存在随机世界之中，对随机现象积累了许多感性的、直观的知识。但是，这些知识却是零散的、无组织的，多数人仍然认为它是一个杂乱无章、难于捉摸的世界。因此，指明每条公理的直观背景是我们的首要任务。不仅如此，我们用一章的篇幅介绍随机世界的直观形象。目的是把随机世界整理成一个较有条理的、富有规律性的世界，为自然公理系统的出现作思想和舆论的准备[②]。

(2) 在引进某组公理后立即用它们推导出一些基本的、常用的结论，并给予证明。其作用有 3 方面：① 使内容成为完整的、系统化的理论；② 具备集合论基础知识的读者能够理解自然公理系统；③ 为专家们指明常见的性质和结论在自然公理系统中所占据的位置。

本书是数学基础性专著。阅读本书只须具备基础性的数学知识，但需要有较强的逻辑思维能力。

---

① 建立和培养因果空间的直观形象是自然公理系统的一项不可缺少的任务，是本书的重要组成部分。

② 第 1 章是具有新观点的科普性内容。

第1章　概率论的现实背景

人们在认识随机世界——自然界和人类社会中各种各样的随机现象时积累了大量的简单观察，发现了一些行之有效、久经考验的实际方法。公理化方法就是从这些素材中提取出一些最本质的属性，进行严密精致的逻辑加工，形成概率论的公理系统。然后以公理系统为基础，建立概率论的理论体系，在新高度上认识随机世界。

本章从最重要的素材中提取出两个原理：

概率论原理Ⅰ：随机事件只有通过它和其他事件之间的联系才能被认识。

概率论原理Ⅱ：在合理划定(或给定)的条件下，随机事件出现的可能性大小能够用[0,1]中惟一的实数 p 表达。

原理Ⅰ指导我们建立自然公理系统的前3组公理。原理Ⅱ用于指导建立自然公理系统的后3组公理。

1.1　研究对象

日月星辰，周而复始；雨雪阴晴，气象万千；江河奔流，汇入大海。生物世界，适者生存，繁衍后代，免遭灭绝。战争与和平，政治与经济，科技与文化，文娱与体育，则是万物之灵——人类的创造。

自然界和人类社会永远处在不断变化和发展之中。反映变化和发展的标志是出现各种各样的现象，或者说，发生形形色色的事情。例如：

(1) 太阳从东方升起(用 A_1 表示这个现象，称为事件 A_1)。

(2) 月亮从太空中消失(事件 A_2)。

(3) 两物体之间存在吸引力(事件 A_3)。

(4) 在标准气压下,水加热到 50℃时沸腾(不可能事件 A_4)。

(5) 明天有雨(事件 A_5)。

(6) 明天阴(事件 A_6)。

(7) 明天晴(事件 A_7)。

(8) 明天的最高气温为 ξ℃(事件 A_8)。

(9) 明天的最低气温为 η℃(事件 A_9)。

(10) 某夫妇的胎儿是男孩(事件 A_{10})。

(11) 某夫妇的胎儿是女孩(事件 A_{11})。

(12) 用第 2 代杂交豌豆培育的某植株开白花(事件 A_{12})。

(13) 用第 2 代杂交豌豆培育的某植株开红花(事件 A_{13})。

事件 A_{12}和 A_{13}取自孟德尔(G. Mendel)做过的著名遗传学实验——豌豆杂交实验。

(14) 罐内放入 n 个球,分别标号为 $1,2,\cdots,n$。任意取出一个球,它是第 i 号球(事件 B_i, $i=1,2,\cdots,n$)。

(15) 100 件产品中有 5 件次品。任意取出 3 件,它们全是次品(事件 A_{15})。

(16) 某日进入某商店的人数恰为 5 千人(事件 A_{16})。

(17) 某日进入某商店的人数不超过 5 千人(事件 A_{17})。

(18) 向桌面抛一枚硬币时正面朝上(事件 A_{18})。

(19) 向桌面抛一枚硬币时反面朝上(事件 A_{19})。

(20) 向桌面抛一枚硬币时硬币直立(事件 A_{20})。

(21) 掷一颗骰子时出现 i 点(事件 C_i, $i=1,2,3,4,5,6$)。

(22) 掷一颗骰子时出现偶数点(事件 A_{22})。

(23) 掷两颗骰子时一颗出现 i 点,另一颗出现 j 点(事件 D_{ij}, $i,j=1,2,3,4,5,6$)。

(24) 掷两颗骰子时出现双 6(事件 A_{24})。

(25) 掷两颗骰子时出现的点数之和大于 8(事件 A_{25})。

诸如此类的事件无穷无尽，不可能一一列举。

所有的科学都研究事件。概率论也是研究事件的科学。它和其他科学的区别在于以下 4 个方面：

(1) 概率论只关心事件是否出现。

(2) 概率论不关心事件本身的其他内容。

(3) 概率论通过事件之间的联系来认识事件。

例如，对于事件 A_7，概率论不是通过 A_7 的具体内容(譬如说，出现"烈日当空"等现象)来认识"明天晴"，而是通过 A_5 和 A_6 等事件不出现来认识"明天晴"。

(4) 如果事件能够出现，则猜测出现的可能性大小。

现象、事情、或事件是否出现受各种各样的条件(包括种种偶然的，不可预测的因素)支配。概率论把它们称为随机现象，或随机事件，用意是强调它们的出现随"机会"而定。事实上，即使像"太阳从东方升起"这样的事件，其出现也依赖于"机会"。因为当你站在其他星球上，或者站在地球的南北极处进行观察，事件 A_1 就未必出现。

现在能够概括出随机事件的一个特点，我们把它概括成如下原理。

概率论原理 I：随机事件只有通过它与其他事件之间的联系才能被认识。

为了用语上的方便，今后把随机事件简称为事件。

1.2 研究任务

《中国大百科全书·数学》中"概率论"条目明确指出："概率论：研究随机现象数量规律的数学分支。"数量规律是一种高度概括的语言。依据 1.1 节的讨论，数量规律应当包括随机现象之间的

联系和随机现象出现的规律。

我们把概率论遵循的一个基本信念写成以下原理。

概率论原理Ⅱ:在合理划定(或给定)的条件下,随机事件出现的可能性大小能够用[0,1]中惟一的实数 p 表达。

在这个原理下,随机现象出现的规律就是我们日常生活中的语言:"在给定的条件下,某随机现象以 $\alpha\%$ 的可能性出现。"或者"在给定的条件下,我们以 $\alpha\%$ 的把握断定(或期待)某随机现象出现。"

用 $\mathscr{C}$ 代表合理划定(或给定)的条件;用 A 代表条件 $\mathscr{C}$ 下涉及的某个随机事件;数值 p 改用符号 $P(A|\mathscr{C})$ 表示,称之为条件 $\mathscr{C}$ 下事件 A 的概率。于是,概率论具有非常单纯的任务:求 $P(A|\mathscr{C})$。

为了完成这项任务,必须把条件 $\mathscr{C}$,事件 A,概率 $P(A|\mathscr{C})$ 抽象成能进行逻辑推理和数学演算的对象。这正是概率论自然公理系统所要做的工作。

在结束本节之前我们给出3点注记。

注1 随机性的本质是什么?遵循的基本信念是否合理?概率论拒绝回答这样的问题。它们是哲学探讨的问题。当然,把在基本信念上所建立的概率理论与现实随机世界进行比较,其吻合程度可以间接地检验基本信念是否合理,是否需要修正。

注2 统计规律是概率论中经常使用的语言。按通常的理解,大量随机现象中蕴含的集体规律性被认为是统计规律。或者更明确地叙述为:"就一次试验而言,看不出什么规律,但是'大数次'地重复这个试验,试验的结果又遵循某些规律,这种规律称之为统计规律,这一类试验称为随机试验,试验所代表的现象称之为随机现象。"[10]

注3 在《现代汉语词典》中规律被解释为"事物之间的内在的必然联系。"在投掷硬币一次的实验中我们有结论:"硬币出现正

面的可能性是50%。”按词典的解释，这个结论不是规律；按注2的解释，这个结论也不是统计规律。但是，这个结论的确是随机现象出现的一个规律。为此，本书认为出现规律就是统计规律。

知识一词比规律包含更多的内涵。它通常解释为“人们在研究和改造世界的实践中获得的认识和经验的总和”。涉及随机现象的知识称为统计知识。在本书中把概率论规定为：**概率论是研究随机现象统计知识的数学分支**。即是说，我们把统计知识等同于“数量规律”，它至少包含两项内容：随机现象之间的联系和统计规律。

1.3 概　　率

为了理解概率论原理Ⅱ，必须探讨概率的含义。这个概念在自然科学和社会科学中已被广泛使用。甚至模式为“明天降雨概率是60%”的天气预报也已为人们普遍接受，似乎概率这一概念不会再有什么异议。“其实不然，仔细考察一下，就会发现各人的理解和作出的解释往往是不同的，这不能不影响到对于概率论、数理统计的看法和应用的范围。这一现象并不是现在才有的，从概率论的发展历史来看，几乎从它诞生时起，就相伴着争论，只是在不同的时期，争论的出发点，争论的重点有所不同，并且影响到学科的内容和发展趋势。”[11]时至今日，还没有一个为所有概率学者都接受的解释。

作者的观点是：每种解释都从某个侧面发掘出概率的本质，重要的是了解在什么样的条件下某种解释才是合理的、可接受的。事实上，多种解释并存不仅使我们正确地、全面地理解“概率是什么”，而且产生许多方面的应用，甚至得到意料之外的、令人惊喜的发现。

下面介绍4种流行的解释。

1.3.1 概率的古典定义

拉普拉斯(P. S. Laplace)首次明确地给出概率的定义(1814年)。他把概率定义为“有利情况”的数目除以“全部可能情况”的总数目。这就是概率的古典定义。在一些场合,古典定义是清楚的、明确的。

例1 向光滑的桌面随意地投掷一枚硬币(实验$\mathscr{E}_1$)。实验的结果不是出现正面,就是出现反面。因此全部可能的情况为

$$〈正面〉,〈反面〉 \tag{1.3.1}$$

用$\mathscr{C}_1$表示确定实验的条件,用$P(〈正面〉|\mathscr{C}_1)$表示〈正面〉出现的概率。由于正面出现只含一种有利情况,按照古典概率的定义得出

$$P(〈正面〉|\mathscr{C}_1)=\frac{1}{2} \tag{1.3.2}$$

同理

$$P(〈反面〉|\mathscr{C}_1)=\frac{1}{2} \tag{1.3.3}$$

例2 向光滑的桌面投掷两枚硬币(实验$\mathscr{E}_2$)。如果用(α,β)表示第1枚硬币出现α面,第2枚出现β面。那么全部可能的情况为

$$〈正,正〉,〈正,反〉,〈反,正〉,〈反,反〉 \tag{1.3.4}$$

至少有一个正面出现包含三个有利情况:〈正,正〉,〈正,反〉,〈反,正〉。所以

$$P(至少有一个正面出现\ |\mathscr{C}_2)=\frac{3}{4} \tag{1.3.5}$$

这里$\mathscr{C}_2$表示确定实验$\mathscr{E}_2$的条件。同理

$$P(出现两个反面\ |\mathscr{C}_2)=\frac{1}{4} \tag{1.3.6}$$

于是得到遗传学中的一个重要比例 3∶1[①]。即

① 参看本节末的讨论。

$$\frac{P(\text{至少有一个正面出现}\mid\mathscr{C}_2)}{P(\text{出现两个反面}\mid\mathscr{C}_2)}=\frac{3}{1}\tag{1.3.7}$$

例 3 向光滑的桌面投掷一颗均匀的骰子(实验 $\mathscr{E}_3$)。这时全部可能的情况为

$$\langle 1\rangle,\langle 2\rangle,\langle 3\rangle,\langle 4\rangle,\langle 5\rangle,\langle 6\rangle\tag{1.3.8}$$

这里$\langle i\rangle$表示出现 i 点。用 $\mathscr{C}_3$ 表示确定实验 $\mathscr{E}_3$ 的条件,应用概率的古典定义得

$$P(\langle 1\rangle\mid\mathscr{C}_3)=P(\langle 2\rangle\mid\mathscr{C}_3)=\cdots=P(\langle 6\rangle\mid\mathscr{C}_3)=\frac{1}{6}\tag{1.3.9}$$

注意,概率的古典定义只在一些特殊场合下才是合理的,可接受的。例如,在例 3 中条件 $\mathscr{C}_3$ 含有骰子均匀、桌面光滑、投掷随意等诸条件。条件 $\mathscr{C}_3$ 的任何变化都将使 $\mathscr{E}_3$ 不遵守概率的古典定义,因而得不到式(1.3.9)。

1.3.2 几何概率

在概率论刚开始发展的时候人们就已经注意到可能情况的总数目和有利情况的数目都可以是无限的。并且出现一些个别的例子,促使对概率进行另一种解释。

这些个别的例子可以一般化为:设 G 是平面(或直线,或空间)中的一个区域,g 是 G 的一个子区域。现在把一个质点随意地投入 G 中(实验 $\mathscr{E}$)。试问质点落在 g 中的概率多大?

问题的回答产生几何概率。用 $\mathscr{C}$ 表示确定实验 $\mathscr{E}$ 的条件,答案是

$$P(\text{质点落入 } g \text{ 中}\mid\mathscr{C})=\frac{\mu(g)}{\mu(G)}\tag{1.3.10}$$

这里视 G,g 为直线、平面、空间中的区域,把 μ 依次地理解为长度、面积和体积。

布丰(de Buffon)对几何概率进行了详细的讨论(1777 年),计

算出许多问题的几何概率。

例 4 （布丰投针实验）平面上画着一些平行线，相邻平行线之间的距离皆等于 a。向平面上随意地投一根长度为 $l(l\leqslant a)$ 的针（实验 $\mathscr{E}_4$），试求此针与任一条平行线相交的概率。

解 解答的关键仍然是找出“全部可能的情况”和“有利的情况”。为此，用 x 表示针的中点到最近一条平行线的距离，用 φ 表示针与平行线的夹角（图 1.1(a)）。显然，满足 $0\leqslant x\leqslant \dfrac{a}{2}$，$0\leqslant\varphi\leqslant\pi$ 的每一对 (φ,x) 确定针落在平面上的一种情况。因此，全部可能的情况组成长方形 G（图 1.1(b)）。

针与平行线相交当且仅当 $x\leqslant\dfrac{l}{2}\sin\varphi$。在 G 中满足这个不等式的 (φ,x) 组成区域 g（图 1.1(b) 中阴影部分）。即事件“针与平行线相交”的有利情况组成区域 g。

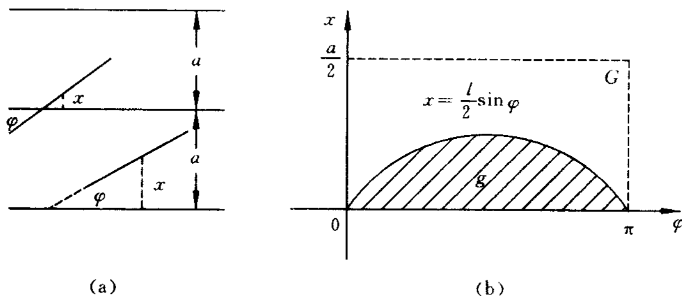


图 1.1

用 $\mathscr{C}_4$ 表示确定实验 $\mathscr{E}_4$ 的条件，并认定概率的几何解释合理，那么由(1.3.10)得出

$$P(\text{针与平行线相交}\mid\mathscr{C}_4)=\frac{g\text{ 的面积}}{G\text{ 的面积}}$$

$$=\frac{2}{\pi a}\int_0^{\pi}\frac{l}{2}\sin\varphi\,\mathrm{d}\varphi=\frac{2l}{\pi a} \qquad (1.3.11)$$

与概率的古典定义一样，几何概率也只在特定的场合下才是合理的、可接受的。

1.3.3 概率的频率解释

很奇怪！人们本能地会接受下述事实：把例1中投掷硬币的实验 $\mathscr{E}_1$ 重复地进行。用 ν_n(正)表示前 n 次投掷中〈正面〉出现的次数，那么 n 很大时有

$$\frac{\nu_n(\text{正})}{n} \approx \frac{1}{2} \tag{1.3.12}$$

称 ν_n(正)$/n$ 为无限投掷序列中事件〈正面〉出现的频率。比较(1.3.2)和(1.3.12)得到概率的频率解释：概率与 n 很大时的频率近似地相等。即

$$P(\langle\text{正面}\rangle|\mathscr{C}_1) \approx \frac{\nu_n(\text{正})}{n} \tag{1.3.13}$$

此式可以用实验进行检验①。表1.1是历史上的实验记录。

表1.1 投掷硬币实验的历史资料

实验者	投掷次数 n	出现正面次数 ν_n(正)	频率 $\frac{\nu_n(\text{正})}{n}$
布　丰	4040	2048	0.5069
德·摩根	4092	2048	0.5005
杰万斯	20480	10379	0.5068
皮尔逊	24000	12012	0.5005
费　勒	10000	4979	0.4979
罗曼诺夫斯基	80640	39699	0.4923

现在把上述想法一般化。给定实验 $\mathscr{E}$，假定它由条件 $\mathscr{C}$ 确定。实验 $\mathscr{E}$ 进行后事件 A 可能出现，也可能不出现。用 $P(A|\mathscr{C})$ 表示

① 对实验序列进行数学描述后概率论将证明(1.3.13)的正确性(参看定理2.7.7和2.7.8)。从而把奇怪的事实变成科学的结论。

事件 A 出现的概率。如果在同样的条件 $\mathscr{C}$ 下重复地进行实验 $\mathscr{E}$，用 $\nu_n(A)$表示前 n 次实验中事件 A 出现的次数，那么

$$P(A|\mathscr{C}) \approx \frac{\nu_n(A)}{n} \tag{1.3.14}$$

并称 $\nu_n(A)/n$ 为事件 A 在该实验序列上的频率。

米泽斯(R. von Mises)把这类实验序列理想化，称之为理想的无限序列。他引进的公理是：理想的无限序列上的频率具有收敛性，其极限值称为概率[12]。在我们的记号下，它是

$$P(A|\mathscr{C}) = \lim_{n\to\infty} \frac{\nu_n(A)}{n} \tag{1.3.15}$$

柯尔莫哥洛夫也是频率解释的支持者。当他把概率定义为一类特殊的测度后立即指明：频率解释是抽象的概率概念和现实世界建立联系的基础[1]。

1.3.4 概率的主观解释

这种解释认为概率存在于人们的主观世界内，它反映人们对某些事物的一种信任程度，是对事物不确定性的一种主观判断。它与个人的知识水平，心理状态诸种因素有关，因此称之为主观概率。

凯恩斯(J. M. Keynes)认为概率是根据其他命题的知识对一个命题作出的一种有理由的信念。对这种信念无法给出数值，只能进行比较[13]。

菲纳特认为每个人提出的概率大小取决于他对指定事件与他将与之比较的事件的判断。萨维奇(L. J. Savage)称之为个人概率[4]。

概率的主观解释适用范围广泛，使用方便，符合人们的习惯。在社会学、经济学等领域中尤其如此。作为例子，我们给出“明天降雨概率是 60%”中概率值 0.6 的各种解释。

(1) 气象台在当前的条件下相信有 60%的可能性降雨，有40%的可能性不降雨。这是概率的主观解释。

(2) 如果设想今天的气象条件能够不断地重复出现，并且 0.6 真实地反映降雨的可能性大小。那么，在不断地重复 n 次预报后大约有 $0.6n$ 次出现明天降雨，大约有 $0.4n$ 次出现明天不降雨。这是概率的频率解释。

(3) 由于(2)中设想的无限预报序列无法实现，于是设计另一个无限序列。把气象台预报"明天降雨概率是 60%"的日子按时间顺序排列，得到一个无限的预报序列。如果气象台的预报方法无重大改进，可以认为它是米泽斯的理想序列。因此，在该序列的前 n 天内大约有 $0.6n$ 天出现降雨，大约有 $0.4n$ 天出现明天不降雨。

如果我们接受解释(1)和(3)，那么能够用统计降雨的天数来判定气象预报的准确性：前 n 天内降雨天数越接近于 $0.6n$，则预报的准确性越高，概率值 0.6 越可信。

概率的古典定义和几何概率在此处失效。

1.3.5 应用初探

概率存在 4 种解释给研究带来许多方便。如果认为两种或更多种的解释是合理的、可信的，那么它们估计的概率值必然一致。否则，至少有一种解释是不合理的。一方面，一致性说明概率是一个科学的概念；另一方面，一致性产生概率论的许多应用。

(1) 投掷硬币的实验 $\mathscr{E}_1$(例 1)中存在概率的古典定义和频率解释。表 1.1 证实两种解释是一致的。

为了确信概率概念的科学性，有兴趣的读者可以对例 2，例 3 或其他实验进行类似的操作，证实一致性是可接受的和合理的。

(2) 作者用下面的实验验证一致性。我们把一颗均匀的骰子在"1 点"，"2 点"和"3 点"的三个面的公共角上灌铅，并配置一个圆柱形的茶叶罐，其高为 9.5cm，底面直径为 7.5cm。把骰子放入

罐内，用力摇动（称为实验 $\mathscr{E}_5$），然后开盖观察骰子出现的点数。

用 $\mathscr{C}_5$ 表示确定实验 $\mathscr{E}_5$ 的条件。显然 $\mathscr{E}_5$ 的全部可能情况是(1.3.8)。事件"i 点"出现的概率值为 $P(\langle i\rangle|\mathscr{C}_5)(i=1,2,\cdots,6)$。与例 3 不同，这 6 个概率值未必相等。应用概率的主观解释我们得出如下结论：

① 由于灌铅的位置使我们当然认为

$$P(\langle i\rangle|\mathscr{C}_5) > P(\langle j\rangle|\mathscr{C}_5) \quad (i = 4,5,6;j = 1,2,3) \tag{1.3.16}$$

成立，并且两端数值的差异是显著的。

② 灌铅时没有要求在角上是"对称的"，所以我们没有把握断定(或期待)以下 3 个概率值

$$P(\langle 1\rangle|\mathscr{C}_5),P(\langle 2\rangle|\mathscr{C}_5),P(\langle 3\rangle|\mathscr{C}_5) \tag{1.3.17}$$

是相等的。

③ 考虑到"4 点"，"5 点"和"6 点"的三个面远离灌铅角，因此灌铅的不对称性对它们的影响都差不多。于是我们又认为应当成立

$$P(\langle 4\rangle|\mathscr{C}_5) = P(\langle 5\rangle|\mathscr{C}_5) = P(\langle 6\rangle|\mathscr{C}_5) \tag{1.3.18}$$

转向概率的频率解释。我们把实验 $\mathscr{E}_5$ 在同样的条件下重复地进行 2000 次，记录下骰子每次出现的点数(有实验记录，略)。用 $\nu_n(i)$ 表示前 n 次实验中 i 点出现的次数，表 1.2 给出该理想的无限序列上的一些频率。

表 1.2 投掷灌铅骰子的实验记录

实验次数 n	⟨1⟩		⟨2⟩		⟨3⟩		⟨4⟩		⟨5⟩		⟨6⟩	
	$\nu_n(1)$	$\frac{\nu_n(1)}{n}$	$\nu_n(2)$	$\frac{\nu_n(2)}{n}$	$\nu_n(3)$	$\frac{\nu_n(3)}{n}$	$\nu_n(4)$	$\frac{\nu_n(4)}{n}$	$\nu_n(5)$	$\frac{\nu_n(5)}{n}$	$\nu_n(6)$	$\frac{\nu_n(6)}{n}$
500	77	0.154	62	0.124	54	0.108	93	0.186	100	0.200	114	0.228
1000	145	0.145	116	0.116	117	0.117	209	0.209	204	0.204	209	0.209
1500	211	0.141	168	0.112	170	0.113	316	0.211	312	0.208	323	0.215
2000	286	0.143	219	0.110	229	0.114	431	0.215	410	0.205	425	0.214

显然，频率解释支持主观解释中的结论①和③，并用等式 $P(\langle 2\rangle|\mathscr{C}_5)=P(\langle 3\rangle|\mathscr{C}_5)$补充结论②。

(3) 把例 4 中实验 $\mathscr{E}_4$ 在同样的条件下重复地进行。用 ν_n 表示前 n 次实验中针与平行线相交的次数，得到事件“针与平行线相交”的频率 ν_n/n。如果认为实验 $\mathscr{E}_4$ 使用几何概率和频率解释都是合理的、可接受的，那么由两种解释的一致性得$\frac{2l}{\pi a}\approx\frac{\nu_n}{n}$。故

$$\pi\approx\frac{2nl}{a\nu_n} \tag{1.3.19}$$

于是人们得出一种用实验数据计算 π 的方法①。反过来，根据 π 的计算值和真值是否一致可以判定两种解释是否合理，是否可信。表 1.3 是历史上的实验记录，它又一次证实概率是一个科学的概念。

表 1.3 投针实验的历史资料(把 a 折算为 1)

实验者	年份	针长	投掷次数	相交次数	π 的实验值
瓦尔夫	1850	0.8	5000	2532	3.1596
史密斯	1855	0.6	3204	1218.5	3.1554
德·摩根	1860	1.0	600	382.5	3.137
福克斯	1884	0.75	1030	489	3.1595
拉查里尼	1901	0.83	3408	1808	3.1415929
里纳	1925	0.5419	2520	859	3.1795

注 拉查里尼在投掷次数不多的前提下得到过于精确的 π 值，统计检验表明他的结果不太可靠[14]。

(4) 孟德尔的豌豆杂交实验

豌豆的花有两种颜色——红花和白花。孟德尔做了如下的工作：

① 提取纯种的红花品种和白花品种。纯种的含义是：这种豌豆品种自花受粉后，后代(不管多少代)总是保持同样的花色。

① 这个方法一般化后发展成蒙特卡罗(Monte-Carlo)方法。

② 用纯种的红花品种和纯种的白花品种做“亲代杂交”[①]，所收获的种子和成长的植株称为子一代。实验结果发现，子一代全部开红花。

③ 用子一代植株自花受粉后所收获的种子和成长的植株称为子二代。实验结果发现，子二代有的开红花，有的开白花。

孟德尔把红花称为显性性状，白花称为隐性性状（由于②），并称它们为一对相对性状。他记录下子二代中红花和白花的植株数后制成表 1.4。

表 1.4　孟德尔豌豆杂交实验结果[②]

编号	相对性状		子二代	子二代中显性植株数		子二代中隐性植株数		子二代中
	显性	隐性	植株总数	数目	%	数目	%	显性植株数：隐性植株数
1	红花	白花	929	705	75.89%	224	24.11%	3.15：1
2	黄子叶	绿子叶	8023	6022	75.06%	2001	24.94%	3.01：1
3	饱满种子	皱瘪种子	7324	5474	74.74%	1850	25.26%	2.96：1
4	成熟豆荚不分节	成熟豆荚分节	1181	822	74.68%	299	25.32%	2.95：1
5	未熟豆荚绿色	未熟豆荚黄色	580	428	73.79%	152	26.21%	2.82：1
6	花腋生	花顶生	858	651	75.87%	207	24.13%	3.14：1
7	高植株	矮植株	1064	787	73.96%	277	26.04%	2.84：1

被遗传学家称之为“神奇比例 3：1”出现在表 1.4 的最后一列。即显性植株数占 3/4，隐性植株数占 1/4。神奇比例是怎样产生的？有些书写道“经过一番冥思苦索，孟德尔终于茅塞顿开。他提出的分离假说完满地说明 3：1 的由来。”[27]

让我们探索“冥思苦索，茅塞顿开”的思维过程。正如遗传学界公认的那样，认为孟德尔成功的原因之一是他比同时代的遗传学家更精通数理科学。我们不妨猜测，孟德尔曾把神奇比例与式

① 即用红花为父，白花为母；或者白花为父，红花为母进行交配。

② 孟德尔同时研究了 7 种相对性状，用了 8 年时间才完成这项实验。

(1.3.7)进行过比较,并制出表 1.5。

表 1.5　豌豆杂交实验和掷两枚硬币实验比较表

相对性状		子二代显性植株数		子二代隐性植株数		显性植株数∶隐性植株数
显　性	隐　性	数目	%	数目	%	
至少有一个正面出现(〈正,正〉,〈正,反〉,〈反,正〉)	出现两个反面(〈反,反〉)	3	75%	1	25%	3∶1
红花	白花	705	75.89%	224	24.11%	3.15∶1
黄子叶	绿子叶	6022	75.06%	2001	24.94%	3.01∶1

比较表 1.5 的最后一列,自然会产生疑问:掷硬币实验出现的 3∶1 与相对性状遗传时出现的 3.01∶1 或 3.15∶1 之间是偶合呢?还是相对性状遗传的机理与掷硬币实验是相似的?孟德尔思想的闪光点在于:他坚信两者之间是相似的,并且依据掷硬币实验描述出相对性状遗传的机理。

于是,表 1.5 启发孟德尔走向伟大的发现,提出遗传学的基因假说:物种的每一种相对性状由一组"基因对"确定,基因对由两个基因组成。基因分为显性和隐性两种。显性性状由显性基因得到;隐性性状由隐性基因得到;当基因对由显性基因和隐性基因配对时表现出显性性状。在表 1.5 中,硬币的正面相当于显性基因,反面相当于隐性基因。

物种如何从上一代获得遗传?孟德尔的遗传学第一定律说:物种交配时确定父本和母本相对性状的基因对按原样分离成两个基因,后代从父本和母本的基因对中各取一个基因,组成自己的基因对。后代获得的基因对决定它具有的相对性状。

基因假说和遗传学第一定律完满地解释了孟德尔的工作①~③。在这项实验中纯种红花的基因对是 CC,纯种白花的基因对是

cc,这里 C,c 分别代表显性和隐性基因。图 1.2 表达出子一代和子二代的遗传过程。

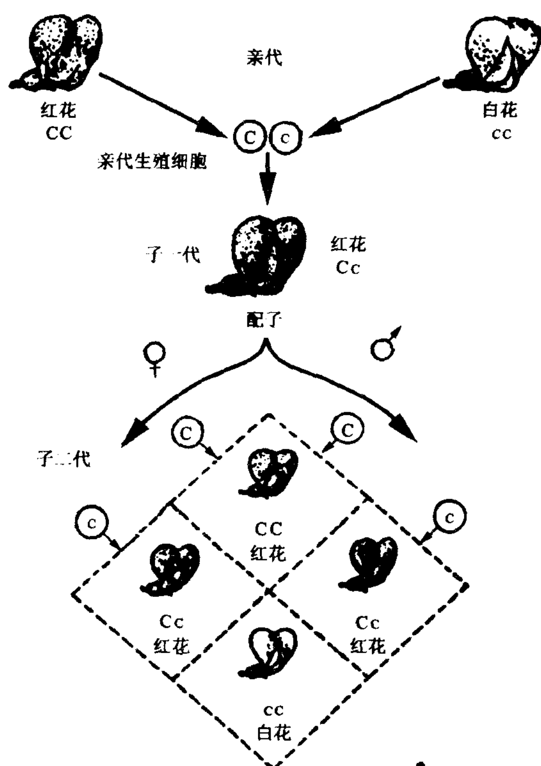


图 1.2 孟德尔的基因假说和遗传学第一定律

遗传学三定律中的另外两条定律也与概率论紧密相关。为了不偏离主题,本书不讨论它们。

第2章　自然公理系统

2.1　概率论元素和6组公理

在概率论的自然公理系统中，随机事件（简称为事件）是惟一不能定义的一个基本概念[①]。它只能像1.1节中那样进行描述。事件被描述清楚，或被确定的标志是：能够判定该事件或者已经出现，或者尚未出现，并且二者必居其一。

习惯上，用英文大写字母$A,B,C\cdots$代表事件。字母A既代表事件A，又含有"事件A出现"或"事件A发生"的意思。自然公理系统包含6组公理。它们是：

Ⅰ　1～3　事件空间公理组

Ⅱ　1～3　原因空间公理组

Ⅲ　1～2　随机试验公理组

Ⅳ　1～3　概率公理组

Ⅴ　1～2　条件概率公理组

Ⅵ　1～6　概率建模公理组

前3组公理反映事件之间的直观联系。它们把随机世界抽象成既形象，又能进行数学演算的因果空间（参看绪论中的概率论与欧氏几何学类比表）。第Ⅳ和第Ⅴ组公理讨论条件与概率之间的内在联系。它们把事件出现的可能性定量化，抽象出概率和条件概率的概念，揭示条件概率和独立性的本质。这5组公理组成概率论的

---

① 定义一个概念时要用到另一个或几个其他的概念。这些"其他的概念"又要用一些概念来定义。最终，总有一批概念是不能给出定义的。称这样的概念为元词。

形式化公理系统。在它的基础上应用逻辑推理和数学演算将演绎出概率论的理论体系。

在概率论的理论体系中,概率是具有某些性质的一些数值(犹如几何学中的距离和力学中的质量)。必须注意,在建立理论体系时只须假定这些数值存在,是给定的。至于它们具体是什么数值,如何获得的,却用不着任何假定。一般理论应用到各个方面,如同几何定理成为物理理论和工程应用的基础一样。

当应用需要确定概率的具体数值时有两种可供使用的方法:(1) 把概率理论和其他理论联系起来;(2) 对观测的数据资料进行统计处理①。统计方法往往是巧妙的。它建筑在概率论的深刻理论上。因此,只有理论向前发展了,才会有更多的确定概率具体数值的方法,才能够看到概率论的广泛应用。

在简化的理想场合,概率的等可能性原理和频率解释是两个久经考验的,能够确定部分概率数值的方法。我们通过第Ⅵ组公理把它们公理化,并建立常用的几类典型的概率空间,对应用和演绎理论体系十分有利。

2.2 第Ⅰ组公理:事件空间公理组

2.2.1 事件空间

概率论不涉及事件的具体内容。孤立地讨论一个事件,事件便成为无法理解的对象。事件只有通过它与其他事件的联系才能被理解,被认识(概率论原理Ⅰ)。因此,概率论要做的第一件事情是把所有的事件归并在一起,并找到事件之间的最基本的、最本质的联系。这是第Ⅰ组公理(它包括3条公理)要完成的任务。

---

① 两种方法的联合使用导致统计方法成为发现、建立和检验"其他理论"的一种重要的科学方法。

定义 2.2.1 用 $\mathscr{U}$ 表示全体事件组成的集合，即

$$\mathscr{U} = \{A \mid A\text{ 为事件}\} \tag{2.2.1}$$

称 $\mathscr{U}$ 为事件空间，如果在 $\mathscr{U}$ 上赋予公理 $I_1 \sim I_3$。

公理 I_1（外延公理） 对于任意的两个事件 A 和 B，如果事件 A 出现时 B 必出现；反之，B 出现时 A 必出现，那么 A 和 B 是同一个事件。

此公理中的事件 A 和 B 被表示为 $A=B$，读为 A 等于 B。

公理 I_2（对立事件公理） 对于任意的事件 A，“A 不出现”也是一个事件。

用 $\overline{A}$ 表示事件“A 不出现”，称为 A 的对立事件，或补事件，简称为补。

在本书中字母 I 永远代表指标集 $\{1,2,\cdots,n\}$ 或 $\{1,2,3,\cdots\}$。前者是有限集，后者是可列集。

公理 I_3（σ 并事件公理） 对于任意的有限个或可列个事件 $A_i, i\in I$，“诸事件 $A_i, i\in I$ 中至少有一个事件出现”也是一个事件。

用 $\bigcup\limits_{i\in I} A_i$ 表示公理 I_3 中的事件，称为 $A_i, i\in I$ 的并事件，简称为并。特别，当 $I=\{1,2,3,\cdots\}$ 时称 $\bigcup\limits_{i\in I} A_i$ 为 σ 并，又记为 $\bigcup\limits_{i=1}^{\infty} A_i$，或 $A_1 \cup A_2 \cup \cdots$。当 $I=\{1,2,\cdots,n\}$ 时称 $\bigcup\limits_{i\in I} A_i$ 为有限并，又记为 $\bigcup\limits_{i=1}^{n} A_i$，或 $A_1 \cup A_2 \cup \cdots \cup A_n$。

假定 $A, B, A_i \in \mathscr{U}(i\in I)$。应用公理 I_2 和 I_3 得 $\overline{A}_i$，$\bigcup\limits_{i\in I} \overline{A}_i$，$\overline{\bigcup\limits_{i\in I} \overline{A}_i}$ 和 $\overline{\overline{A} \cup B}$ 皆为事件。为了更细致地刻画相等关系和深刻理解补事件和并事件，我们再引进一种关系和两种产生新事件的方法。

定义 2.2.2 设 A 和 B 是事件，如果事件 A 出现时事件 B 必

出现，则称 A 是 B 的子事件，或称 A 包含于 B，或称 B 包含 A。记为

$$A \subset B, \quad 或 B \supset A \tag{2.2.2}$$

定义 2.2.3 称 $\overline{\bigcup_{i \in I} \overline{A}_i}$ 为诸事件 $A_i, i \in I$ 的交事件，简称为交，简记为 $\bigcap_{i \in I} A_i$。

与并事件类似，当 $I = \{1,2,3,\cdots\}$ 时称 $\bigcap_{i \in I} A_i$ 为 σ 交，又记为 $\bigcap_{i=1}^{\infty} A_i$，或 $A_1 \cap A_2 \cap A_3 \cap \cdots$，或 $A_1A_2A_3\cdots$；当 $I = \{1,2,\cdots,n\}$ 时称 $\bigcap_{i \in I} A_i$ 为有限交，又记为 $\bigcap_{i=1}^{n} A_i$，或 $A_1 \cap A_2 \cap \cdots \cap A_n$，或 $A_1A_2\cdots A_n$。

定义 2.2.4 称 $\overline{\overline{A} \cup B}$ 为事件 A 和 B 的差事件，简称为差，简记为 $A \backslash B$。

引理 2.2.1 假定 $A, B \in \mathscr{U}$，那么 $A = B$ 当且仅当 $A \subset B$ 且 $B \subset A$。

证明 结论是公理 I_1 和定义 2.2.2 的直接推论。 **证毕**

引理 2.2.2 对任意的 $A \in \mathscr{U}$ 成立 $\overline{\overline{A}} = A$。

证明 注意到事件的属性是：它或者出现，或者不出现，并且二者必居其一。因此，事件"$\overline{A}$ 不出现"就是事件"A 出现"。使用对立事件的记号和公理 I_1 即得等式 $\overline{\overline{A}} = A$。 **证毕**

引理 2.2.3 交事件 $\bigcap_{i \in I} A_i$ 表示"诸事件 $A_i, i \in I$ 皆出现"这一事件。

证明 用 $\Leftrightarrow$ 表示当且仅当。由定义 2.2.3 得出

$$\bigcap_{i \in I} A_i \text{ 出现} \overset{\text{公理 } \mathrm{I}_2}{\Longleftrightarrow} \bigcup_{i \in I} \overline{A}_i \text{ 不出现}$$

$$\overset{\text{公理 } \mathrm{I}_3}{\Longleftrightarrow} \text{诸事件 } \overline{A}_i, i \in I \text{ 皆不出现}$$

$\xleftrightarrow{\text{引理 2.2.2}}$ 诸事件 $A_i, i \in I$ 皆出现 **证毕**

引理 2.2.4 差事件 $A \backslash B$ 表示"事件 A 出现,但 B 不出现"这一事件。

证明 由定义 2.2.4 得出

$A \backslash B$ 出现 $\xleftrightarrow{\text{公理 } I_2}$ $\overline{A} \cup B$ 不出现

$\xleftrightarrow{\text{公理 } I_3}$ $\overline{A}$ 和 B 皆不出现

$\xleftrightarrow{\text{引理 2.2.2}}$ A 出现,但 B 不出现 **证毕**

于是,公理 $I_1 \sim I_3$,定义 2.2.2~2.2.4 在事件空间 $\mathscr{U}$ 中规定了 6 种关系"$=, \subset, \ , \cup, \cap, \backslash$"。应用这 6 种关系能够方便地表达出事件之间的其他关系,从而使我们获得关于 $\mathscr{U}$ 的丰富的感性知识。例如

$\overline{A} \cap \overline{B}$ 表示"事件 A 和 B 皆不出现"这一事件;

$A \cup B \subset C$ 表示 $A \cup B$ 是 C 的子事件。或者说,当事件 A 和 B 中至少有一个出现时事件 C 必出现;

$(A \cap B) \backslash C = D$ 表示"A 和 B 同时出现,但 C 不出现"是事件 D;

$\bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$ 表示"存在正整数 n,使得诸事件 $A_i, i \geqslant n$ 皆出现"这一事件。事实上,由公理 I_3 推出存在某正整数 n 使得 $\bigcap_{i=n}^{\infty} A_i$ 出现。再应用引理 2.2.3 即得结论。

同理,$\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$ 表示事件"诸 $A_i, i \geqslant 1$ 中至少有无穷多个事件出现"。

请读者给出更多的例子。因为熟悉大量的这类事件将增加对事件空间 $\mathscr{U}$ 的理解,也是掌握概率论的关键的一步。

2.2.2 必然事件和不可能事件

定义 2.2.5 (1) 称 $\mathscr{U}$ 中事件 X 为必然事件,如果对任意的

$A \in \mathscr{U}$ 成立 $A \subset X$。

(2) 称 $\mathscr{U}$ 中事件 $\varnothing$ 为不可能事件，如果对任意的 $A \in \mathscr{U}$ 成立 $\varnothing \subset A$。

引理 2.2.5 下列 4 个结论相互等价

(1) $A \subset B$；(2) $\overline{B} \subset \overline{A}$；(3) $A \cup B = B$；(4) $A \cap B = A$

证明 由于"事件 A 出现时 B 一定出现"等价于"事件 B 不出现时 A 一定不出现"，后者可以叙述为"$\overline{B}$ 出现时 $\overline{A}$ 一定出现"。得证(1)和(2)相互等价。

假定(3)成立。公理 I_3 蕴含"A 出现时 $A \cup B$ 一定出现"，应用(3) 即得 B 一定出现，故 $A \subset B$。反之，假定(1)成立。由于 A 出现时 B 一定出现，因此 A 和 B 至少有一个出现时 B 也一定出现，故 $A \cup B \subset B$。公理 I_3 还蕴含 $B \subset A \cup B$。联合两个包含式即得(3)(引理 2.2.1)。得证(1)和(3)等价。类似地可证(1)和(4)等价。

证毕

定理 2.2.1 事件空间 $\mathscr{U}$ 中存在必然事件 X 和不可能事件 $\varnothing$。并且它们是惟一的。

证明 由于 $\mathscr{U}$ 非空，任取 $B \in \mathscr{U}$。由公理 I_2 推出 $\overline{B} \in \mathscr{U}$。令

$$X = B \cup \overline{B} \tag{2.2.3}$$

由公理 I_3 得出 $X \in \mathscr{U}$。对于任意的 $A \in \mathscr{U}$，由于 B 和 $\overline{B}$ 中必有一个事件出现，故 A 出现时 $B \cup \overline{B}$ 一定出现，从而 $A \subset X$。得证 X 是必然事件。令

$$\varnothing = \overline{X} \tag{2.2.4}$$

由公理 I_2 得出 $\varnothing \in \mathscr{U}$。对任意的事件 A，我们有 $\overline{A} \subset X$。由引理 2.2.5推出 $\overline{X} \subset \overline{\overline{A}}$。应用引理 2.2.2 即得 $\varnothing \subset A$，得证 $\varnothing$ 是不可能事件。

以下证惟一性。设 X_1 是另一个必然事件，则 $X_1 \subset X$ 且 $X \subset X_1$。由引理 2.2.1 推出 $X_1 = X$，得证必然事件的惟一性。同理可证不可能事件的惟一性。

证毕

定义 2.2.5 的直观背景：由(2.2.3)得出 X 是事件空间中一定会出现的事件；由(2.2.4)得出 $\varnothing$ 是一定不出现的事件。

2.2.3 运算规则

事件空间 $\mathscr{U}$ 中 4 种产生新事件的方法“补、并、交、差”之间存在相互联系。由此导致 4 种方法成为事件的 4 种运算，它们之间的联系成为运算法则。

定理 2.2.2(并、交的 σ 交换-结合律)　设指标集 I 有两个子集 I_1 和 I_2(允许它们相交)，并且 $I_1 \cup I_2 = I$[①]。那么成立

$$\bigcup_{i \in I} A_i = \left[\bigcup_{j \in I_1} A_j\right] \cup \left[\bigcup_{k \in I_2} A_k\right] \tag{2.2.5}$$

$$\bigcap_{i \in I} A_i = \left[\bigcap_{j \in I_1} A_j\right] \cap \left[\bigcap_{k \in I_2} A_k\right] \tag{2.2.6}$$

证明　假定事件 $\bigcup\limits_{i \in I} A_i$ 出现。由公理 I_3 知诸 $A_i, i \in I$ 中至少有一个事件出现。不妨设 A_r 出现。$r \in I$ 蕴含 $r \in I_1$ 和 $r \in I_2$ 中至少有一个成立，故事件 $\bigcup\limits_{j \in I_1} A_j$ 和 $\bigcup\limits_{k \in I_2} A_k$ 中至少有一个出现。得证

$$\bigcup_{i \in I} A_i \subset \left[\bigcup_{j \in I_1} A_j\right] \cup \left[\bigcup_{k \in I_2} A_k\right]$$

反之，假定(2.2.5)右方的事件出现。由公理 I_3 知事件 $\bigcup\limits_{j \in I_1} A_j$ 和 $\bigcup\limits_{k \in I_2} A_k$ 中至少有一个出现。不妨设 $\bigcup\limits_{j \in I_1} A_j$ 出现。再次应用公理 I_3 得出，至少存在一个 $r \in I_1$ 使事件 A_r 出现。注意到 $r \in I_1$ 时有 $r \in I$，所以事件 $\bigcup\limits_{i \in I} A_i$ 出现。得证

$$\left[\bigcup_{j \in I_1} A_j\right] \cup \left[\bigcup_{k \in I_2} A_k\right] \subset \bigcup_{i \in I} A_i$$

---

① 注意，这里的 $\cup$ 是集合论中的并运算。事件的补、并、交、差和集合论中补、并、交、差有同样的称呼和记号决不是偶然的，可参看 2.3 节。

联合上述两个包含式即得(2.2.5)。同理可证(2.2.6)。 **证毕**

推论 并、交运算具有性质：

(1) 幂等律

$$A \cup A = A;\ A \cap A = A$$

(2) 交换律

$$A \cup B = B \cup A;\ A \cap B = B \cap A$$

(3) 结合律

$$(A \cup B) \cup C = A \cup (B \cup C);\ (A \cap B) \cap C = A \cap (B \cap C)$$

定理 2.2.3(并和交的 σ 分配律) 设 $B, A_i \in \mathscr{U}(i \in I)$，那么成立

$$\left[\bigcup_{i \in I} A_i\right] \cap B = \bigcup_{i \in I} [A_i \cap B] \tag{2.2.7}$$

$$\left[\bigcap_{i \in I} A_i\right] \cup B = \bigcap_{i \in I} [A_i \cup B] \tag{2.2.8}$$

证明 假定(2.2.7)左方事件出现。由引理 2.2.3 推出事件 $\bigcup_{i \in I} A_i$ 和 B 皆出现。应用公理 I_3 得出，至少存在一个 $j \in I$ 使得事件 A_j 和 B 皆出现，即 $A_j \cap B$ 出现。从而(2.2.7)右方事件出现。得证

$$\left[\bigcup_{i \in I} A_i\right] \cap B \subset \bigcup_{i \in I} [A_i \cap B]$$

反之，假定(2.2.7)右方事件出现。则至少存在一个 $j \in I$ 使 $A_j \cap B$ 出现(公理 I_3)，故事件 A_j 和 B 皆出现(引理 2.2.3)，从而 $\bigcup_{i \in I} A_i$ 和 B 皆出现(公理 I_3)。得证

$$\bigcup_{i \in I} [A_i \cap B] \subset \left[\bigcup_{i \in I} A_i\right] \cap B$$

联合两个包含式即得(2.2.7)。同理可证(2.2.8)。 **证毕**

定理 2.2.4 补、并、交、差运算具有性质

(1) $A \backslash B = A \cap \overline{B}$ (2.2.9)

(2) 德·摩根(de Morgan)律

$$\overline{\bigcup_{i\in I} A_i} = \bigcap_{i\in I} \overline{A}_i \tag{2.2.10}$$

$$\overline{\bigcap_{i\in I} A_i} = \bigcup_{i\in I} \overline{A}_i \tag{2.2.11}$$

证明 (1) 由引理 2.2.4 知 $A\backslash B$ 是事件"A 出现,但 B 不出现"。应用公理 I_2 得出后者等价于"A 和 $\overline{B}$ 皆出现"。再应用引理 2.2.3即得(2.2.9)。

(2) 由定义 2.2.3 知

$$\bigcap_{i\in I} A_i = \overline{\bigcup_{i\in I} \overline{A}_i} \tag{2.2.12}$$

用 $\overline{A}_i$ 代替其中的 A_i,应用引理 2.2.2 即得(2.2.10)。在(2.2.12)两边取对立事件,应用引理 2.2.2 即得(2.2.11)。 **证毕**

2.2.4 事件域和事件 σ 域

定义 2.2.6 设 $\mathscr{F}$ 是部分事件组成的非空集合。称 $\mathscr{F}$ 为事件域,如果它满足条件(1)和(2):

(1) 关于补封闭:若 $A\in\mathscr{F}$,则 $\overline{A}\in\mathscr{F}$。

(2) 关于并封闭:若 $A,B\in\mathscr{F}$,则 $A\cup B\in\mathscr{F}$。

称 $\mathscr{F}$ 是事件 σ 域,如果它满足条件(1)和(3):

(3) 关于 σ 并封闭:若对任意的 $i\in I$ 有 $A_i\in\mathscr{F}$,则 $\bigcup_{i\in I} A_i \in \mathscr{F}$。

显然,事件空间 $\mathscr{U}$ 本身是事件域,也是事件 σ 域。

定理 2.2.5 (1) 事件域包含必然事件 X 和不可能事件 $\varnothing$,并且关于有限并、有限交和差运算也封闭。

(2) 事件 σ 域是事件域。它关于可列交也封闭。

证明 (1) 由于 $\mathscr{F}$ 非空,故存在 $B\in\mathscr{F}$。由定义 2.2.6 中条件(1)和(2)得出必然事件 $X=B\cup\overline{B}\in\mathscr{F}$。由此又得出不可能事件 $\varnothing=\overline{X}\in\mathscr{F}$。

假定 $A_i \in \mathscr{F}(i=1,2,\cdots,n)$。我们有 $\bigcup_{i=1}^{3} A_i=[A_1 \cup A_2] \cup A_3 \in \mathscr{F}$。应用归纳法可证 $\mathscr{F}$ 关于有限并封闭。

在(2.2.12)中取 $I=\{1,2,\cdots,n\}$，并应用有限并的封闭性和定义 2.2.6 中条件(1)得出 $\mathscr{F}$ 关于有限交封闭。

由(2.2.9)和有限交的封闭性得出 $\mathscr{F}$ 关于差的封闭性。

(2) 当取 $I=\{1,2\}$ 时定义 2.2.6 中条件(3)成为条件(2)，故事件 σ 域是事件域。由(2.2.12)，应用条件(1)和(3)得出 $\mathscr{F}$ 关于可列交封闭。　　**证毕**

定义 2.2.7　设 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 是两个事件 σ 域(或事件域)。如果成立集合的包含式 $\mathscr{F}_2 \subset \mathscr{F}_1$，则称 $\mathscr{F}_2$ 是 $\mathscr{F}_1$ 的子事件 σ 域(或子事件域)。

显然，任何事件 σ 域(或事件域)皆是事件空间 $\mathscr{U}$ 的子事件 σ 域(或事件域)；$\mathscr{F}_0=\{\varnothing, X\}$ 是任何事件 σ 域(或事件域)的子事件 σ 域(或事件域)。不妨称 $\mathscr{F}_0$ 为最小事件 σ 域或最小事件域。

2.2.5 注记

(1) 应用必然事件和不可能事件，有限并与有限交可以作为可列并与可列交的特例。实现方法为

$$\bigcup_{i=1}^{n} A_i=A_1 \cup A_2 \cup \cdots \cup A_n \cup \varnothing \cup \varnothing \cup \cdots \tag{2.2.13}$$

$$\bigcap_{i=1}^{n} A_i = A_1 A_2 \cdots A_n X X \cdots \tag{2.2.14}$$

(2) 设 A 和 B 是两个事件。称 A 和 B 是不相容的，如果 $AB=\varnothing$。一个非常有用的结论是：事件的并可以转化为不相容事件的并。事实上，对于并事件 $\bigcup_{i\in I} A_i$，令

$$B_1 = A_1, B_2 = A_2 \backslash B_1, \cdots, B_n = A_n \backslash \left(\bigcup_{i=1}^{n-1} B_i \right), \cdots \tag{2.2.15}$$

容易证明，诸 $B_i (i \in I)$ 互不相容，并且

$$\bigcup_{i\in I} A_i = \bigcup_{i\in I} B_i \tag{2.2.16}$$

(3) 在事件空间 $\mathscr{U}$ 中还可以引进其他的运算。例如，可以用

$$A\Delta B \stackrel{\text{def}}{=\!=} (A\backslash B) \cup (B\backslash A) \tag{2.2.17}$$

定义运算 Δ，称为"对称差"运算。

(4) 设 J 是任意的指标集，它的元素可以是有限个、可列个或不可列个。本书中指标集I 永远属于前两种情形。不可列指标集有：

实数集 $\mathbf{R}$；区间$[a,b]$；矩形$\{(x,y)\,|\,a\leqslant x\leqslant b, c\leqslant y\leqslant d\}$

等等。这里 a,b,c,d 是实数，并且 $a<b, c<d$。

假定 $A_\alpha\in\mathscr{U}(\alpha\in J)$。在概率论中偶尔会使用语句"诸事件 $A_\alpha,\alpha\in J$ 中至少有一个出现"和"诸事件 $A_\alpha,\alpha\in J$ 皆出现"，并分别记为

$$\bigcup_{\alpha\in J} A_\alpha;\quad \bigcap_{\alpha\in J} A_\alpha \tag{2.2.18}$$

注意，这两个记号仅当 J 是有限或可列集时才代表事件。当 J 是不可列集时它们既不是 $\mathscr{U}$ 中的运算，也不能保证是 $\mathscr{U}$ 中的事件。它们仅仅是两个记号。

(5) 在 K 公理系统和自然公理系统中可列无限和有限处于同等的地位，不可列无限才是"真正的无限"。为了强调这点，带 σ 字样的语言中的 σ 通常不省略。

2.3 第Ⅱ组公理：原因空间公理组

2.3.1 问题和方法

事件空间 $\mathscr{U}$ 是全体随机事件形成的一个大集合。它是抽象的，因为我们无法列举出它的全部元素，甚至不知道它的势有多大[①]；它又是具体的，因为我们能够列举出许许多多的事件。问题

① 集合的势反映集合所含元素的多少。特别，有限集合的势等于它所含元素的个数。

是:如何清楚地、形象地认识事件空间$\mathscr{U}$。

类似的问题在科学中曾多次遇到。例如

(1) 世界上存在各种各样的物质。试问:如何认识所有的物质?

经过许多人的探索后德谟克利特(Demokritos)宣告:万物的本原是一种叫原子的东西。大小不等,形状不同的原子在位置和次序方面以不同的排列构成了万物。

为了认识所有的物质,人们找到了大小不等,形状不同的原子——一百多种元素。然后再研究元素如何构成万物——各种各样的混合物和化合物[①]。

(2) 人类和动植物繁衍后代。后代的个体之间既有相同的性状,又有不相同的性状。试问:如何认识各种各样的遗传性状?

孟德尔宣告:在物种中存在一种叫做"基因对"的东西,它由两个基因组成。一种基因对代表一种遗传性状。后代从父亲和母亲处各获得一个基因形成自己的基因对。后代获得的基因对确定了它的遗传性状。

(3) 在现实世界中存在各种各样的图形。试问:如何研究图形?

几何学宣称,人们能够设想出最小的"图形"——点。全体想象中的点形成欧氏空间。线和面是由一些特殊点组成的集合。应用线和面可以在空间中围成各种各样的图形。

(4) 在现实世界中存在各种各样的随机事件。试问:如何研究随机事件?

在上述思想的启迪下概率论的自然公理系统认为,人们能够设想出最小的"随机事件"——原因点。全体想象中的原因点形成原因空间。随机事件是由一些特殊的原因点组成的集合。应用随

---

① 所有的物质形成一个大集合。同样,我们也不知道这个集合的势有多大。甚至,世上有多少种元素我们也不知道。

机事件可以产生原因空间(确切地说,是因果空间)中各种各样的"图形"——随机试验,概率空间等。

最小单位——元素和基因对是现实世界中的物质,应用科学方法可以观测到它们;点和原因点则是人们抽象思维的产物,只有通过人类的思维才能理解它们。

2.3.2 原因点的直观背景

一个"最小的随机事件"被设想为所有事件的一次虚构的实现。它可以用如下的方法表达。用 1 和 0 分别代表出现和不出现。当事件 A 出现时认为 A 对应数值 1,否则对应数值 0。于是,"所有事件的一次虚构的实现"是一个定义在 $\mathscr{U}$ 上取值 0 或 1 的一个函数

$$u(A)=\begin{cases}1 & A\text{ 出现}\\ 0 & A\text{ 不出现}\end{cases} \tag{2.3.1}$$

我们称这样的函数为原因点[①],简记为 u。

原因点 u 代表事件空间 $\mathscr{U}$ 的一种状态:凡使 $u(B)=1$ 的事件 B 皆出现,凡使 $u(C)=0$ 的事件 C 皆不出现。称原因点 u 伪出现,如果事件空间 $\mathscr{U}$ 处于 u 所代表的状态中[②]。

---

① 原因点的正式定义将在第(3)段中给出。借此机会申明,标有直观背景的章、节和段落中不能定义自然公理系统中的任何概念和符号。这样的章、节和段落有:本段落,2.4 节、2.7 节和 2.9 节的第(1)段,3.1 节和第 1 章。后者是第 Ⅰ 组,第 Ⅳ 组和第 Ⅵ 组公理的直观背景。

在这些章、节和段落中,我们通过对现实随机世界的分析,抽象出随机现象最本质的东西,形成概念的原形,指导自然公理系统定义这些概念和引进公理。因此在建立公理系统时它们是不可缺少的内容。

由于这部分内容是抽象的公理系统与现实世界之间联系的桥梁,因此它们也是理解自然公理系统和进行概率论应用时必须具备的知识。

② 原因点 u 不是事件。为了判定事件空间 $\mathscr{U}$ 是否处于 u 所代表的状态中需要知道所有的事件中哪些已经出现,哪些尚未出现。显然,这是办不到的事情。即是说,我们无法判定"现象"u 是否已经出现,故 u 不是事件。

之所以称 u 为原因点，是因为用逆向思维的方式我们可以说：由于原因点 u 伪出现，所以事件 B（它使 $u(B)=1$）出现，事件 C（它使 $u(C)=0$）不出现。

注意，不是每个定义在 $\mathscr{U}$ 上取值 0 或 1 的函数都是原因点。例如，给定事件 D，使 $u(D)=u(\overline{D})=1$ 的函数 $u(A)$，$A\in\mathscr{U}$ 就不是原因点。因为由公理 I_2 知事件 D 和 $\overline{D}$ 皆出现的状态是不存在的。事实上，应用第 I 组公理容易推出原因点具有下列 4 条性质：

(1) $u(X)=1$；$u(\varnothing)=0$。

(2) $u(\overline{A})=1-u(A)$，对任意的 $A\in\mathscr{U}$。

(3) $u\left(\bigcup_{i\in I}A_i\right)=\sup_{i\in I}\{u(A_i)\}$，对任意的 $A_i\in\mathscr{U}(i\in I)$。

(4) $u\left(\bigcap_{i\in I}A_i\right)=\inf_{i\in I}\{u(A_i)\}$，对任意的 $A_i\in\mathscr{U}(i\in I)$。

并且性质(1)、(4)可用(2)、(3)推出①。

现在固定 $\mathscr{U}$ 中的一个事件 B，定义一个原因点的集合

$$\{u\,|\,u\text{ 是满足 }u(B)=1\text{ 的原因点}\}$$

直观上，该集合代表一组原因。如果该集合中某个原因点伪出现，则称这组原因出现。于是，当这组原因出现时事件 B 一定出现，当它不出现时事件 B 一定不出现，其他的事件则在遵守性质(2)和(3)的前提下可以自由地选择出现或不出现。因此，我们可以把这组原因和它的后果——事件 B 等同。即

$$B=\{u\,|\,u\text{ 是满足 }u(B)=1\text{ 的原因点}\}\qquad(2.3.2)$$

并且**事件** B 出现当且仅当**集合** B 中的某个原因点伪出现。

至此，我们得出结论——事件是由一些原因点组成的集合。特别，当 $B=\varnothing$ 时(2.3.2)的右方是空集，因此不可能事件又称为空集，或空事件。当 $B=X$ 时(2.3.2)成为

① 去掉符号和运算的概率意义后，在代数学中 $\mathscr{U}$ 是布尔(G. Boole)σ 代数，集合 $\{0,1\}$ 赋予适当运算后是布尔代数。满足性质(2)和(3)的函数 $u(A)$，$A\in\mathscr{U}$ 称为 $\mathscr{U}$ 到 $\{0,1\}$ 上的 σ 同态映射[15,16]。

$$X = \{u \mid u \text{ 是原因点}\} \tag{2.3.3}$$

它是所有原因点组成的集合。所以必然事件又可以称为原因空间。

小结：我们引进概念的先后顺序为：事件⇒事件空间⇒原因点⇒事件是原因点的集合⇒必然事件是原因空间。

引进原因空间的意义在于：事件成为原因空间的子集。即将证明，事件在 $\mathscr{U}$ 中的 6 种关系——相等、包含、补、并、交、差等同于它作为 X 的子集时在集合论中的相应关系。二元组$(X,\mathscr{U})$成为集合论中的可测空间。于是，概率论找到了最有效的研究工具——集合论方法。

文氏(J. Venn)图给出集合论中 6 种关系的直观形象。从而也给出事件空间中 6 种关系的直观形象(图 2.1)。

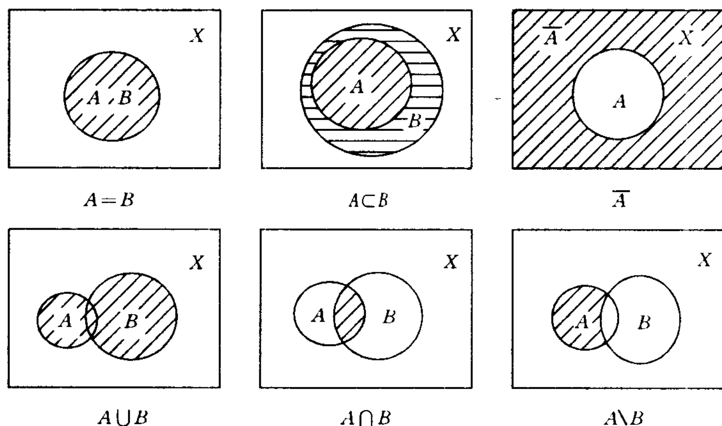


图 2.1 事件的 6 种关系

2.3.3 原因空间公理组

定义 2.3.1 设 $u(A), A \in \mathscr{U}$ 是定义在事件空间 $\mathscr{U}$ 上取值 0 或 1 的函数。如果它具有上段中的性质(1)～(4)，则称它为原因点，简记为 u。

公理 II_1（原因点公理） 原因点 u 存在。它代表事件空间 $\mathscr{U}$ 所处的一种状态：凡使 $u(B)=1$ 的事件 B 皆出现，凡使 $u(C)=0$ 的事件 C 皆不出现。称原因点 u 伪出现，如果 $\mathscr{U}$ 处于 u 所代表的状态中。

公理 II_2（伪出现公理） 称全体原因点组成的集合 X 为原因空间。原因空间中有且只有一个原因点伪出现。

公理 II_3（事件公理） 事件 B 是原因空间的子集。它是

$$B = \{u \mid u \text{ 是满足 } u(B) = 1 \text{ 的原因点}\}$$

并且 B 出现是指 B 中的某个原因点伪出现。

定义 2.3.2 称二元组 $(X,\mathscr{U})$ 为因果空间，如果 $\mathscr{U}$ 是事件空间，X 是原因空间。

定理 2.3.1 在因果空间 $(X,\mathscr{U})$ 中，$\mathscr{U}$ 中事件及其运算与 X 中集合及其运算之间呈现出表 2.1 所表达的一致性。

证明 第 I 组公理建立表 2.1 中基本要素之间的对应。下证基本关系之间的对应。

(1) 设 $B\subset C$ 是事件的包含式。对任意的原因点 $u(A)$，$A\in\mathscr{U}$，应用 $C=B\cup C$ 和原因点的性质(3)，我们有

$$u(C) = u(B \cup C) = \sup\{u(B),u(C)\}$$

假定原因点 $u^*\in B$，由公理 II_3 知 $u^*(B)=1$。由上式推出 $u^*(C)=1$。再次应用公理 II_3 知 $u^*\in C$。得证 $B\subset C$ 也是集合论中的包含式。

反之，设 $B\subset C$ 是集合论中的包含式，并且 B 和 C 是事件。假定事件 B 出现。由公理 II_3 知 B 中的某个原因点 u^* 伪出现。由集合包含式 $B\subset C$ 得 $u^*\in B$ 推出 $u^*\in C$，故 C 中有原因点 u^* 伪出现。再次应用公理 II_3 得出事件 C 出现。得证 $B\subset C$ 是概率论中的包含式。

(2) 由于在集合论中也有引理 2.2.1，应用这个引理和已证的(1)推出相等关系在概率论和集合论中是一致的。

表 2.1　概率论语言与集合论语言对照表

因果空间 $(X,\mathscr{U})$			
	学科 符号	集合论	概率论
		原因空间 X	事件空间 $\mathscr{U}$
基本要素	u	原因点	缺
	G	任意子集	缺
	A	特殊子集(一组原因)	事件
	$\varnothing$	空集	不可能事件
	X	原因空间	必然事件
	$\mathscr{U}$	一个特殊的 σ 域	事件空间
	$(X,\mathscr{U})$	一个特殊的可测空间	因果空间
在下面的关系中,A,B,A_i,A_α 皆为事件,I 为有限或可列指标集,J 为不可列指标集			
基本关系	$u\in A$	原因点 u 是 A 的元素	u 伪出现导致事件 A 出现
	$A=B$	A 和 B 由同一些原因点组成	A 出现时 B 必出现,并且 B 出现时 A 必出现
	$A\subset B$	A 中的原因点皆含于 B 中	事件 A 出现时 B 必出现
	$\overline{A}$	由不属于 A 的所有原因点组成的集合	对立事件:“A 不出现”
	$\bigcup_{i\in I}A_i$	由所有 $A_i(i\in I)$中的原因点组成的集合	并事件:“诸 $A_i,i\in I$ 中至少有一个事件出现”
	$\bigcup_{\alpha\in J}A_\alpha$	由所有 $A_\alpha(\alpha\in J)$中的原因点组成的集合	缺
	$\bigcap_{i\in I}A_i$	由同时属于所有 $A_i(i\in I)$中的原因点组成的集合	交事件:“诸事件 $A_i,i\in I$ 皆出现”
	$\bigcap_{\alpha\in J}A_\alpha$	由同时属于所有 $A_\alpha(\alpha\in J)$中的原因点组成的集合	缺
	$A\backslash B$	由属于 A,但不属于 B 的原因点组成的集合	差事件:“A 出现,但 B 不出现”
	$A\cap B=\varnothing$	集合 A 和 B 不相交	事件 A 和 B 不相容

(3) 设 B 和 $\overline{B}$ 是对立事件。由公理 II_3 知 B 有表达式(2.3.2),$\overline{B}$ 有表达式

$$\overline{B}=\{u\,|\,u\text{ 是满足 }u(\overline{B})=1\text{ 的原因点}\}\qquad(2.3.4)$$

注意到原因点 $u(A),A\in\mathscr{U}$ 是只取 0 和 1 两个值的函数,应用原

因点的性质(2)得出

$$u^* \in B \Leftrightarrow u^*(B) = 1 \Leftrightarrow u^*(\overline{B}) = 0 \Leftrightarrow u^* \overline{\in} \overline{B}$$

得证 $\overline{B}=X\backslash B$,即 $\overline{B}$ 是 B 在集合论中的补。

反之,设 B 是原因空间 X 的子集,并且是事件。B 在集合论中的补集用 $X\backslash B$ 表示。注意到原因点是只取 0 和 1 两个值的函数,由(2.3.2)推出

$$X\backslash B = \{u \mid u \text{ 是满足 } u(B) = 0 \text{ 的原因点}\} \qquad (2.3.5)$$

用 $\overline{B}$ 表示 B 的对立事件。原因点的性质(2)含有:u 满足条件 $u(B)=0$当且仅当 $u(\overline{B})=1$。故(2.3.5)成为

$$X\backslash B = \{u \mid u \text{ 是满足条件 } u(\overline{B}) = 1 \text{ 的原因点}\}$$

由公理 II_3 知 $\overline{B}=X\backslash B$,即 B 在集合论中的补也是它在概率论中的补。

(4) 设 $B_i(i \in I)$ 是事件,$\bigcup_{i\in I} B_i$ 是并事件。由公理 II_3 知,它们分别是 X 的子集

$$B_i=\{u \mid u \text{ 是满足 } u(B_i)=1 \text{ 的原因点}\} \quad (i\in I) \qquad (2.3.6)$$

和

$$\bigcup_{i\in I} B_i = \left\{u \mid u \text{ 是满足 } u\left(\bigcup_{i\in I} B_i\right) = 1 \text{ 的原因点}\right\} \qquad (2.3.7)$$

由原因点 u 的性质(3)推出:$u\left(\bigcup_{i\in I} B_i\right) = 1$ 当且仅当至少存在某个 B_i 使 $u(B_i) = 1$。因此,$\bigcup_{i\in I} B_i$ 是由所有 $B_i, i \in I$ 中原因点组成的集合,得证 $\bigcup_{i\in I} B_i$ 也是诸 $B_i, i \in I$ 在集合论中的并。

反之,假定对任意的 $i\in I$,B_i 是 X 的子集,并且是事件。由公理 II_3 知 B_i 有形式(2.3.6)。用 B^* 表示诸 $B_i, i\in I$ 在集合论中的并。由集合论中并的定义得出

$$B^*=\{u \mid u \text{ 是原因点,并且存在某个 } B_i \text{ 使 } u(B_i)=1\}$$

用 $\bigcup_{i\in I} B_i$ 表示诸 $B_i, i \in I$ 在事件空间中的并事件。应用(2.3.7)后

关于原因点的结论得出

$$B^* = \left\{u \mid u \text{ 是满足 } u\left(\bigcup_{i\in I} B_i\right) = 1 \text{ 的原因点}\right\}$$

于是，由公理 II_3 知 B^* 是事件，并且 $B^* = \bigcup_{i\in I} B_i$。得证并关系在概率论和集合论中是一致的。

(5) 由于德·摩根律在概率论和集合论中皆成立。应用(2.2.12)和已证的(3)，(4)即得交的一致性。

(6) 由于在概率论和集合论中皆有 $A\backslash B = A\cap\overline{B}$。应用此式和已证的(3)，(5)即得差的一致性。 **证毕**

2.3.4 注记

(1) 因果空间的哲学解释

$$B = \{u \mid u \text{ 是满足 } u(B) = 1 \text{ 的原因点}\} \tag{2.3.2}$$

的右方是原因空间 X 的一个子集，它代表一组原因。(2.3.2)的左方是事件空间 $\mathscr{U}$ 的一个元素，它代表这组原因产生的后果。于是，(2.3.2)的右方和左方构成一组因果关系，并被形象地表达为

$$\text{后果} = \text{原因组} \tag{2.3.8}$$

并且，后果出现当且仅当原因组中的某个原因发生(即伪出现)。特别，令 $B=X$ 时它成为

$$\text{必然事件} = \text{全部原因} \tag{2.3.9}$$

因此，(2.3.2)是日常生活中的语言"由因导果，执果索因"的数学表达式。

在现实中人们关于因果律的思维方式包含两项内容：

① 根据原因推断后果，或者根据后果寻找原因。

② 当原因改变时推断后果的相应变化。

第①项内容的数学抽象是(2.3.2)。第②项内容是(2.3.2)的数学演算，即原因的改变被抽象成(2.3.2)右方的集合运算，后果的变化被抽象成左方的事件运算。于是，自然公理系统把因果律抽

象成能进行数学演算的定量化的哲学规律。

(2) 表 2.1 指明，事件的不可列并 $\bigcup_{\alpha\in J} A_\alpha$ 和不可列交 $\bigcap_{\alpha\in J} A_\alpha$ 在概率论中是无意义的运算。但在集合论意义下它们是有意义的运算，并且产生 X 的新子集。至于这个新子集是否为事件则要小心判断。一般说来它不是事件，但在特殊情形下(例如对一切 $\alpha\in J$ 皆有 $A_\alpha=X$)它可以是事件。

(3) 公理 II_1 保证原因点存在。公理 II_3 保证当 $B\neq\varnothing$ 时(2.3.2)的右方不是空集。这两个保证不仅是信念，而且是可以证明的数学结论。事实上，它们是布尔 σ 代数中著名的斯通(M. H. Stone)表现定理的内容[15,16]。

之所以用公理的形式肯定这两个保证，是因为它们很直观。我们既不想涉及过多的代数学知识，更不想用抽象的代数定理掩盖概率论的直观性和形象性。

(4) 下节用到集合论中划分的概念。这里给出简单的介绍。

定义 2.3.3 设 E 是集合，J 是指标集，$\Pi=\{\pi_\alpha|\alpha\in J\}$ 是由 E 的子集 π_α 组成的集系。如果 Π 满足如下 3 个条件：

(1) $\pi_\alpha\neq\varnothing$，对任意的 $\alpha\in J$；

(2) $\pi_\alpha\cap\pi_\beta=\varnothing$，对任意的 $\alpha,\beta\in J$ 且 $\alpha\neq\beta$；

(3) $\bigcup_{\alpha\in J}\pi_\alpha=E$。

则称 Π 为 E 的一个划分，且称 π_α 为划分块。如果 Π 满足条件(2)和(3)，则称 Π 为 E 的广义划分。

定义 2.3.4 设 $\Pi_1=\{\pi_\alpha^{(1)}|\alpha\in J_1\}$，$\Pi_2=\{\pi_\beta^{(2)}|\beta\in J_2\}$ 是 E 的两个划分。如果对任意的 $\pi_\beta^{(2)}\in\Pi_2$，存在 Π_1 中划分块 $\pi_r^{(1)}$ 使得

$$\pi_\beta^{(2)}\subset\pi_r^{(1)} \tag{2.3.10}$$

则称划分 Π_2 是 Π_1 的加细，或 Π_1 是 Π_2 的加粗。

例 1 设 $A_i(i\in I)$ 是非空事件，诸事件互不相容(即对任意不同的 $i,j\in I$ 有 $A_iA_j=\varnothing$)，并且 $\bigcup_{i\in I}A_i=X$。显然 $\{A_i|i\in I\}$ 是

原因空间 X 的一个划分。

例 2 设 $A_\alpha(\alpha \in J, J$ 是不可列集)是事件,诸事件互不相容。如果诸集合 $A_\alpha, \alpha \in J$ 的并 $\bigcup_{\alpha \in J} A_\alpha = X$,那么$\{A_\alpha | \alpha \in J\}$是原因空间 X 的一个广义划分。如果补充假定:诸 $A_\alpha(\alpha \in J)$ 皆是非空事件。那么$\{A_\alpha | \alpha \in J\}$是 X 的一个划分。

例 3 设 $\mathbf{R}=(-\infty,+\infty)$。显然

$$\Pi_1 = \{\mathbf{R}\} \tag{2.3.11}$$

$$\Pi_2 = \{(n, n+2] | n \text{ 为奇数}\} \tag{2.3.12}$$

$$\Pi_3 = \{(n, n+2] | n \text{ 为偶数}\} \tag{2.3.13}$$

$$\Pi_4 = \{(n, n+1] | n \text{ 为整数}\} \tag{2.3.14}$$

$$\Pi_5 = \{\{x\} | x \in \mathbf{R}\} \tag{2.3.15}$$

都是 $\mathbf{R}$ 的划分。并且

(1) Π_1 是最粗的划分。即是说,不存在比 Π_1 更粗的划分。或者说,$\mathbf{R}$ 的任何划分都是 Π_1 的加细。

(2) Π_5 是最细的划分。即是说,不存在比 Π_5 更细的划分。或者说,$\mathbf{R}$ 的任何划分都是 Π_5 的加粗。

(3) Π_4 是 Π_2 的加细,也是 Π_3 的加细。

(4) Π_2 和 Π_3 不能比较粗细。

引理 2.3.1 设 $\Pi_1=\{\pi_\alpha^{(1)} | \alpha \in J_1\}$,$\Pi_2=\{\pi_\beta^{(2)} | \beta \in J_2\}$是集合 E 的两个划分,并且 Π_2 是 Π_1 的加细。那么对任意的 $\pi_\alpha^{(1)} \in \Pi_1$,存在 J_2 的子集 J_α 使得

$$\pi_\alpha^{(1)} = \bigcup_{\beta \in J_\alpha} \pi_\beta^{(2)} \tag{2.3.16}$$

引理 2.3.2 设 $\Pi_1=\{\pi_\alpha^{(1)} | \alpha \in J_1\}$,$\Pi_2=\{\pi_\beta^{(2)} | \beta \in J_2\}$是集合 E 的两个划分。那么

$$\Pi_1 * \Pi_2 \overset{\text{def}}{=\!=} \{\pi_\alpha^{(1)} \cap \pi_\beta^{(2)} | \alpha \in J_1, \beta \in J_2\} \tag{2.3.17}$$

是 E 的一个广义划分。

定义 2.3.5 从 $\Pi_1 * \Pi_2$ 中删除空集后的划分称为 Π_1 和 Π_2 的交叉划分，记为 $\Pi_1 \circ \Pi_2$。如果 $\Pi_1 * \Pi_2$ 不含空集，则称 $\Pi_1 \circ \Pi_2$ $(=\Pi_1 * \Pi_2)$ 为真交叉划分。

例 4 对于例 3 中的划分，我们给出如下 4 个结论：

(1) 设 Π 是 $\mathbf{R}$ 的任意划分，那么

$$\Pi_1 \circ \Pi = \Pi_1 * \Pi = \Pi \tag{2.3.18}$$

是 Π_1 和 Π 的真交叉划分。

(2) 设 Π 是 $\mathbf{R}$ 的任意划分，但 $\Pi \neq \Pi_1$。显然，$\Pi_5 * \Pi$ 是 $\mathbf{R}$ 的广义划分，但不是划分。因此

$$\Pi_5 \circ \Pi = \Pi_5 \tag{2.3.19}$$

是 Π_5 和 Π 的交叉划分，但不是真交叉划分。

(3) $\Pi_2 * \Pi_3 = \{(n,n+2] \cap (m,m+2] \mid n$ 为奇数，m 为偶数$\}$ 是 $\mathbf{R}$ 的广义划分，但不是划分。故

$$\Pi_2 \circ \Pi_3 = \Pi_4 \tag{2.3.20}$$

是 Π_2 和 Π_3 的交叉划分，但不是真交叉划分。

(4) $\Pi_3 * \Pi_4 = \{(n,n+2] \cap (m,m+1] \mid n$ 为偶数，m 为整数$\}$ 是广义划分，但不是划分。故

$$\Pi_3 \circ \Pi_4 = \Pi_4 \tag{2.3.21}$$

是 Π_3 和 Π_4 的交叉划分，但不是真交叉划分。

例 5 给定欧氏平面 $\mathbf{R}^2 = \{(x,y) \mid x,y \in \mathbf{R}\}$。对任意固定的实数 x 和 y，令

$$(x,\mathbf{R}) = \{(x,z) \mid z \in \mathbf{R}\} \tag{2.3.22}$$

$$(\mathbf{R},y) = \{(z,y) \mid z \in \mathbf{R}\} \tag{2.3.23}$$

显然

$$\Pi_1 = \{(x,\mathbf{R}) \mid x \in \mathbf{R}\} \tag{2.3.24}$$

$$\Pi_2 = \{(\mathbf{R},y) \mid y \in \mathbf{R}\} \tag{2.3.25}$$

是 $\mathbf{R}^2$ 的两个划分。

$$\Pi_1 \circ \Pi_2 = \Pi_1 * \Pi_2 = \mathbf{R}^2 \tag{2.3.26}$$

是 Π_1 和 Π_2 的真交叉划分。

例 6 假定

$$\Omega_1 = \{\omega_\alpha^{(1)} \mid \alpha \in J_1\} \tag{2.3.27}$$

$$\Omega_2 = \{\omega_\beta^{(2)} \mid \beta \in J_2\} \tag{2.3.28}$$

是原因空间 X 的两个划分。其中 J_1, J_2 是两个指标集，其元素可以是有限个，可列个或不可列个。那么

$$\Omega_1 \circ \Omega_2 = \{\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \mid \alpha \in J_1, \beta \in J_2, \omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \neq \varnothing\} \tag{2.3.29}$$

是 Ω_1 和 Ω_2 的交叉划分。

再假定对任意的 $\alpha \in J_1, \beta \in J_2$ 有 $\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \neq \varnothing$。那么

$$\Omega_1 \circ \Omega_2 = \{\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \mid \alpha \in J_1, \beta \in J_2\} \tag{2.3.30}$$

是 Ω_1 和 Ω_2 的真交叉划分。

2.4 第Ⅲ组公理：随机试验公理组

2.4.1 直观背景

前两组公理把随机世界抽象成因果空间 $(X, \mathscr{U})$。遗憾的是，我们既不能把 X，又不能把 $\mathscr{U}$ 中的元素一一列举①。通常只能研究一些孤立化、理想化后的局部，以及局部之间的关系。

经验告诉我们，所谓的局部是我们关心的某些事件。其个数可以是有限的，可列的，或不可列的。用 $\mathscr{F}$ 表示这个局部，并设

$$\mathscr{F} = \{A, A_i (i \in I), B, C \cdots\}$$

直观上，由这些事件用补、并、交、差产生的事件也是我们关心的事件。即是说

---

① 用 E 表示所有化学元素的集合，用 $\mathscr{M}$ 表示所有物质的集合。那么 $(E, \mathscr{M})$ 相当于 $(X, \mathscr{U})$。在研究物质时人们也不能把 E 或 $\mathscr{M}$ 中的元素一一列举。

$$\bar{A},\ \bigcup_{i\in I}A_i,\ \bigcap_{i\in I}A_i, B\backslash C$$

也属于 $\mathscr{F}$。因此 $\mathscr{F}$ 关于补、并、交、差运算封闭①。简言之，$\mathscr{F}$ 是事件 σ 域。

为了研究这个局部，我们首先把研究对象从 $(X,\mathscr{U})$ 缩小为 $(X,\mathscr{F})$。其中 $\mathscr{F}$ 的元素能够一一列举，或者概括地描述。

其次，用 $\mathscr{F}$ 代替 $\mathscr{U}$ 进行类似于 2.3 节的讨论。令

$$\Omega=\{\omega|\omega \text{ 是定义在 } \mathscr{F} \text{ 上取值 0 或 1 的函数}\} \quad (2.4.1)$$

其中的 ω 还满足上节第(1)段中的性质(1)～(4)(当然，那里的 $\mathscr{U}$ 皆要改为 $\mathscr{F}$)。直观上，ω 代表 $\mathscr{F}$ 处于这样的一种状态：凡使 $\omega(B)=1$ 的事件 B 皆出现，凡使 $\omega(C)=0$ 的事件 C 皆不出现。

同样的理由得出，$\mathscr{F}$ 中的事件皆是 Ω 的子集。即 $B\in\mathscr{F}$ 时有

$$B=\{\omega|\omega\in\Omega, \text{并且 } \omega(B)=1\} \quad (2.4.2)$$

第三，讨论 Ω 和原因空间 X 的关系。比较(2.3.2)和(2.4.1)得出

$$\begin{aligned}\omega=&\{u|u \text{ 是满足 } u(B)\\ &=\omega(B)(B\in\mathscr{F}) \text{ 的原因点}\}\end{aligned} \quad (2.4.3)$$

$$X=\bigcup_{\omega\in\Omega}\omega \quad (2.4.4)$$

假定 ω_1 和 ω_2 是 Ω 中不同的元素，由(2.4.3)得出 ω_1 和 ω_2 不含相同的原因点。于是得出，Ω 是原因空间 X 的一个划分。

第四，事件 B 现在有 3 种表达方法。在 $\mathscr{U}$ 中是字母 B，在 X 中是(2.3.2)的右方，在 Ω 中是(2.4.2)的右方。即

$$B=\{\omega|\omega\in B\}=\{u|u\in B\} \quad (2.4.5)$$

其中前一个等式是(2.4.2)的简写；$B=\{u|u\in B\}$ 是(2.3.2)的简写；后一个等式是集合运算

$$\{\omega|\omega\in B\}\overset{\text{def}}{=\!=}\bigcup_{\omega\in B}\omega=\{u|u\in B\} \quad (2.4.6)$$

---

① 第Ⅱ组公理是这个直观事实的数学抽象。

的简写。

通过上述讨论，我们把研究对象缩小为$(\Omega,\mathscr{F})$。这个转变的重要性在于：Ω 和 $\mathscr{F}$ 中的元素能够一一列举，或者概括地描述。

2.4.2 随机试验公理组

定义 2.4.1 如果 $\Omega=\{\omega_\alpha|\alpha\in J\}$是原因空间 X 的一个划分，则称 Ω 为样本空间，ω_α 为样本点。称样本点 ω_α 伪出现，如果 ω_α 中的某个原因点伪出现。

定义 2.4.2 设 B 是事件，Ω 是样本空间。如果 B 可以表示成 Ω 的子集

$$B=\{\omega_\alpha|\alpha\in J_1\} \tag{2.4.7}$$

则称 B 是 Ω 上的事件。其中 J_1 是定义 2.4.1 中指标集 J 的子集，(2.4.7)按(2.4.6)理解。

定义 2.4.3 称$(\Omega,\mathscr{F})$为随机试验，如果 Ω 是样本空间，$\mathscr{F}$ 是 Ω 上事件的非空集合，并且遵守公理 $Ⅲ_1$ 和 $Ⅲ_2$。

公理 $Ⅲ_1$(补封闭公理) 对 $\mathscr{F}$ 中任意的事件 A 成立 $\overline{A}\in\mathscr{F}$。

公理 $Ⅲ_2$(σ 并封闭公理) 对 $\mathscr{F}$ 中任意的有限个，或可列个事件 $A_i,i\in I$ 成立 $\bigcup\limits_{i\in I}A_i\in\mathscr{F}$。

由定义 2.2.6 看出 $\mathscr{F}$ 是事件 σ 域。但它与一般的事件 σ 域不同，$\mathscr{F}$ 中的事件皆是 Ω 上的事件。今后，称这类特殊的 $\mathscr{F}$ 为 Ω 上的事件 σ 域。于是，定义 2.4.3 可以简述为

定义 2.4.3′ 称$(\Omega,\mathscr{F})$为随机试验，如果 Ω 是样本空间，$\mathscr{F}$ 是 Ω 上的事件 σ 域。

为了用语上的方便，随机试验有时简称为试验①。首先把几个

---

① 在概率论的书籍中试验和实验往往不加区分，并且未能给出确切的定义。在自然公理系统中随机试验成为一个有确切定义的数学名词。

明显的事实写成

引理 2.4.1 (1) 对任意的事件 σ 域 $\mathscr{F}$,$(X,\mathscr{F})$是随机试验。

(2) 给定事件 σ 域 $\mathscr{F}$,Ω 用(2.4.1)规定,则$(\Omega,\mathscr{F})$是随机试验。

(3) 设$(\Omega,\mathscr{F})$是随机试验,划分 Ω^* 是 Ω 的加细,那么$(\Omega^*,\mathscr{F})$是随机试验。

这个引理说明一个非常重要的事实:样本空间仅仅是研究事件 σ 域的一件工具,在一定的意义下可以自由地选择。今后,称(2)中的$(\Omega,\mathscr{F})$为 $\mathscr{F}$ 的生存试验。

定理 2.4.1 设$(\Omega,\mathscr{F})$是随机试验,那么

(1) Ω 中有且仅有一个样本点伪出现。

(2) $\mathscr{F}$ 中事件 B 出现当且仅当 B 中的某个样本点伪出现。

证明 (1) 设 $\Omega=\{\omega_\alpha|\alpha\in J\}$是样本空间。由公理 II_2 知原因空间 X 中有且仅有一个原因点 u 伪出现。由于 Ω 是 X 的划分,故这个原因点属于且仅属于一个样本点 ω_α。由定义 2.4.1 得出 Ω 中有且仅有这个样本点 ω_α 伪出现。

(2) 设 B 有形式(2.4.7)。假定事件 B 出现,那么 B 中的某个原因点伪出现(公理 II_3)。由(2.4.7)推出这个原因点属于 B 中某个样本点 ω_α,因此 ω_α 伪出现。反之,假定 B 中某个样本点 ω_α 伪出现。那么存在某个 $u\in\omega_\alpha$ 使原因点 u 伪出现(定义 2.4.1)。即 B 中原因点 u 伪出现,故 B 出现(公理 II_3)。 **证毕**

类似于定理 2.3.1 的讨论得到

定理 2.4.2 在随机试验$(\Omega,\mathscr{F})$中,$\mathscr{F}$ 中事件及其运算和 Ω 上的集合及其运算呈现出表 2.2 所表达的一致性。

表 2.2 随机试验中概率论术语与集合论术语对照表

随机试验$(\Omega,\mathscr{F})$			
学科 \ 符号		集合论术语	概率论术语
		样本空间Ω	Ω上事件σ域$\mathscr{F}$
基本要素	ω	样本点	可能是$\mathscr{F}$中事件,也可能不是
	F	一般子集	可能是$\mathscr{F}$中事件,也可能不是
	A	特殊子集(一组原因)	$\mathscr{F}$中事件
	$\varnothing$	空集	不可能事件
	Ω	样本空间	必然事件
	$\mathscr{F}$	特殊的σ域	Ω上事件σ域
	$(\Omega,\mathscr{F})$	特殊的可测空间	随机试验
在基本关系对应中A,B,A_i,A_j皆为$\mathscr{F}$中事件,I是有限或可列指标集,J是不可列指标集			
基本关系	$\omega\in A$	ω是集合A的元素	ω伪出现导致A出现
	$A\subset B$	A内的每个样本点皆属于B	A出现时B必出现
	$A=B$	A与B含有相同的样本点	A与B或同时出现,或同时不出现
	$\bar{A}$	由不属于A的所有样本点组成的集合	对立事件"A不出现"
	$\bigcup\limits_{i\in I}A_i$	由所有$A_i,i\in I$中样本点组成的集合	并事件"诸$A_i,i\in I$中至少有一个事件出现"
	$\bigcup\limits_{j\in J}A_j$	由所有$A_j,j\in J$中样本点组成的集合	缺
	$\bigcap\limits_{i\in I}A_i$	由同时属于诸$A_i,i\in I$中样本点组成的集合	交事件"诸事件$A_i,i\in I$皆出现"
	$\bigcap\limits_{j\in J}A_j$	由同时属于诸$A_j,j\in J$中样本点组成的集合	缺
	$A\backslash B$	由属于A,但不属于B的样本点组成的集合	差事件"A出现,但B不出现"

2.4.3 子随机试验

定义 2.4.4 设$(\Omega_i,\mathscr{F}_i)(i=1,2)$是两个随机试验。如果

$\mathscr{F}_2 \subset \mathscr{F}_1$，则称$(\Omega_2, \mathscr{F}_2)$是($(\Omega_1, \mathscr{F}_1)$的子随机试验，简称为子试验。特别，$\Omega_1 = \Omega_2$时称为同因子试验。

显然，因果空间$(X, \mathscr{U})$本身是一个随机试验。任何随机试验皆是它的子试验。

(1) $\Omega_0 = \{X\}$是原因空间的最粗划分，$\mathscr{F}_0 = \{\varnothing, X\}$是最小的事件$\sigma$域。显然，$(\Omega_0, \mathscr{F}_0)$是随机试验，而且是任何试验的子试验。称$(\Omega_0, \mathscr{F}_0)$为退化试验。

(2) 设A是任意的事件。显然$\Omega = \{A, \overline{A}\}$是原因空间$X$的一个划分，$\mathscr{F} = \{\varnothing, A, \overline{A}, \Omega\}$是$\Omega$上的事件$\sigma$域。因此$(\Omega, \mathscr{F})$是随机试验，称为用事件$A$生成的试验，或伯努利(J. Bernoulli)试验。

在绪论中我们把因果空间比喻为欧氏空间，把随机试验比喻为欧氏空间中的图形。因此，熟悉大量的“图形”——随机试验，对培养因果空间的空间形象和学习概率论都是不可缺少的事情。从现在开始直至2.6节末我们都是做这项工作。

2.4.4 随机试验建模(I)：列举法

概率论学科包括两类工作。一类是用公理系统建立理论体系。使用的方法是逻辑推理和数学演算。对于这类工作，只要前提成立，推演出的结论肯定正确。

另一类工作是把现实世界中人们关心的随机现象抽象成理论体系中的研究对象(我们称这部分工作为概率论建模，简称为建模)；然后进行前一类工作中的推演，得到各种各样的结论；最后用这些结论表达出所关心的随机现象中蕴含的统计规律。

建模工作是一种反映与被反映的关系。人们不能谈论建模工作是否正确，只能谈论是否合理，是否成功。下面给出判定建模合理性的3条准则。

准则1 模型保留实际问题的某些本质，适合使用者的要求，因而人们接受这个模型。

准则 2 从一个实际问题中允许抽象出多个不同的模型。原因是建模时各自强调和扬弃了实际问题的某些方面。

准则 3 模型反映实际问题的一种理想化面貌。它和实际问题之间的符合程度需要用实践检验。

建模工作在应用中普遍存在。当几何学被应用时人们有时把太阳、地球等天体视为质点，有时视为球体或椭球体；建筑在地面的一道围墙有时认为是平面上的曲线，有时认为是带形。这样的工作属于几何学的建模。由于几何学的建模过于灵活而又变化无常，更由于人们"天生地"具备这类建模本领，所以几何学拒绝将建模内容作为自身的一个部分。概率论则不同，人们对模型(即概率论中的研究对象)和建模方法都比较陌生，因此必须用大量的、典型的例子熟悉研究对象和建模方法。

例 1 向桌面投掷一枚硬币(实验 $\mathscr{E}_1$)，试求 $\mathscr{E}_1$ 产生的随机试验。

解 实验 $\mathscr{E}_1$ 进行后不是出现正面，就是出现反面。故两个事件"正面"和"反面"是对立事件，

$$\Omega_1 = \{正面,反面\} \tag{2.4.8}$$

是必然事件。因而 Ω_1 是原因空间 X 的一个划分，即 Ω_1 是样本空间①。令

$$\mathscr{F}_1 = \{\varnothing,\langle正面\rangle,\langle反面\rangle,\Omega_1\} \tag{2.4.9}$$

显然 $\mathscr{F}_1$ 是 Ω_1 上的事件 σ 域。于是，$(\Omega_1,\mathscr{F}_1)$是 $\mathscr{E}_1$ 产生的随机试验。

答案$(\Omega_1,\mathscr{F}_1)$有明显的直观背景。当人们询问："向桌面投掷一枚硬币时将出现什么样的结果？"一些人可能回答："不知道"；另

---

① 能够列举出的样本点，特别是应用问题中的样本点几乎都是事件。可能是这个原因，柯尔莫哥洛夫把米泽斯引进的样本点称为基本事件[1]。于是人们又称样本空间为基本事件空间(在[1]中称为基本集合)。即便如此，对于随机试验$(\Omega,\mathscr{F})$，基本事件也可以不是 $\mathscr{F}$ 中的事件。参看例 6 和例 7。

一些人可能回答："不是出现正面，就是出现反面"；还有一些人会回答："出现的结果是且只能是 Ω_1 中的一个基本事件，我们关心的所有事件组成事件 σ 域 $\mathscr{F}_1$"。

在生活中三种回答都是合理的，可接受的。但是，只有第三种回答才是科学的。因为随机试验 $(\Omega_1,\mathscr{F}_1)$ 全面地、准确地反映实验 $\mathscr{E}_1$ 的结果；更因为任何实验的结果都可以用随机试验进行全面的、准确的描述。

例 2 向桌面投掷一枚硬币（实验 $\mathscr{E}_2$），试求 $\mathscr{E}_2$ 产生的随机试验。

解 实验 $\mathscr{E}_2$ 进行后可能出现"正面"，可能出现"反面"，可能出现"站立"。并且认定只能是这三种情况中的一种。故

$$\Omega_2 = \{正面,反面,站立\} \tag{2.4.10}$$

是必然事件，也是原因空间 X 的一个划分。令

$$\mathscr{F}_2 = \Omega_2 \text{ 的所有子集组成的集合} \tag{2.4.11}$$

易证 $\mathscr{F}_2$ 是 Ω_2 上的事件 σ 域。显然，$(\Omega_2,\mathscr{F}_2)$ 是 $\mathscr{E}_2$ 产生的随机试验。

生活中的"实验"一词有时是明确的，有时是含糊的。所谓明确的，是指实验遵守的条件是人人共同认为的"不言而喻"的条件。所谓含糊的，是指人们不能把实验遵守的条件一个不漏地罗列出来。因此总能根据漏掉的条件人为地把一个实验变成为两个或几个。当前，向桌面投掷硬币的实验就是这样的一个例子。当我们遵守"不言而喻"的条件时它是实验 $\mathscr{E}_1$；当桌面有窄裂缝时它是实验 $\mathscr{E}_2$。

科学的一个任务就是消除这类"含糊性"。由于随机试验准确地规定了实验的部分条件，在进行理论研究时我们用试验 $(\Omega_i,\mathscr{F}_i)$ 作为刻画实验 $\mathscr{E}_i$ 的部分条件[①]；在建模工作中则用"不言

---

① 即是说，实验 $\mathscr{E}_i$ 产生的随机试验是 $(\Omega_i,\mathscr{F}_i)$。2.7 节引进的概率空间 $(\Omega_i,\mathscr{F}_i,P_i)$ 将全面地、准确地刻画实验 $\mathscr{E}_i$ 的全部条件。

而喻”的条件作为实验进行的条件，人们将依据“不言而喻”的条件建立实验产生的试验。

例 3 在一无所知的情况下预测某胎儿的性别（实验 $\mathscr{E}_3$），试求 $\mathscr{E}_3$ 产生的随机试验。

解 由于胎儿的性别不是男，就是女。故

$$\Omega_3 = \{男, 女\} \tag{2.4.12}$$

是 $\mathscr{E}_3$ 产生的样本空间。令

$$\mathscr{F}_3 = \{\varnothing, \{男\}, \{女\}, \Omega_3\} \tag{2.4.13}$$

它是 Ω_3 上的事件 σ 域。显然，$(\Omega_3, \mathscr{F}_3)$ 是 $\mathscr{E}_3$ 产生的随机试验。

例 4 预测杂交豌豆第二代的某植株的花色（实验 $\mathscr{E}_4$），试求 $\mathscr{E}_4$ 产生的随机试验。

解 由 1.3 节得知植株不是开红花，就是开白花。故

$$\Omega_4 = \{红花, 白花\} \tag{2.4.14}$$

是 $\mathscr{E}_4$ 产生的样本空间。令

$$\mathscr{F}_4 = \{\varnothing, \{红花\}, \{白花\}, \Omega_4\} \tag{2.4.15}$$

它是 Ω_4 上的事件 σ 域。显然，$(\Omega_4, \mathscr{F}_4)$ 是 $\mathscr{E}_4$ 产生的随机试验。

例 5 向桌面投掷一枚骰子（实验 $\mathscr{E}_5$），试求 $\mathscr{E}_5$ 产生的随机试验。

解 由于骰子出现的点数只能是 1，2，3，4，5，6 中的一个。故

$$\Omega_5 = \{1, 2, 3, 4, 5, 6\} \tag{2.4.16}$$

是必然事件，也是原因空间 X 的一个划分。即 Ω_5 是 $\mathscr{E}_5$ 产生的样本空间。令

$$\mathscr{F}_5 = \Omega_5 \text{ 的所有子集组成的集合} \tag{2.4.17}$$

它是 Ω_5 上的事件 σ 域。因此 $(\Omega_5, \mathscr{F}_5)$ 是 $\mathscr{E}_5$ 产生的随机试验。

例 6 在上例的实验中我们只关心出现点数的奇偶性（实验 $\mathscr{E}_6$），试求 $\mathscr{E}_6$ 产生的随机试验。

解法 1 采用上例的记号。子集 $\{1,3,5\}$ 和 $\{2,4,6\}$ 分别是事件“出现奇数”和“出现偶数”。令

$$\mathscr{F}_6 = \{\varnothing, \{1,3,5\}, \{2,4,6\}, \Omega_5\} \tag{2.4.18}$$

它是 Ω_5 上的事件 σ 域，并且是实验 $\mathscr{E}_6$ 关心的全部事件。故 $(\Omega_5, \mathscr{F}_6)$ 是 $\mathscr{E}_6$ 产生的随机试验。

解法 2 用 ω_o 和 ω_e 分别表示事件“出现奇数”和“出现偶数”。由于投掷骰子的结果不是奇数就是偶数，故

$$\Omega_6 = \{\omega_o, \omega_e\} \tag{2.4.19}$$

是必然事件，也是原因空间的一个划分。即 Ω_6 是 $\mathscr{E}_6$ 产生的一个样本空间。令

$$\mathscr{F}_6 = \{\varnothing, \langle\omega_o\rangle, \langle\omega_e\rangle, \Omega_6\} \tag{2.4.20}$$

它是 Ω_6 上的事件 σ 域。因此 $(\Omega_6, \mathscr{F}_6)$ 是 $\mathscr{E}_6$ 产生的随机试验。

例 7 在例 5 的实验中把 5 和 6 视为大点，把 1,2,3,4 视为小点。如果我们只关心出现大点和小点（实验 $\mathscr{E}_7$），试求 $\mathscr{E}_7$ 产生的随机试验。

解 用 ω_b 和 ω_s 分别表示事件“出现大点”和“出现小点”。由于投掷骰子不是出现大点就是出现小点。故

$$\Omega_7 = \{\omega_b, \omega_s\} \tag{2.4.21}$$

是必然事件，也是原因空间 X 的一个划分。令

$$\mathscr{F}_7 = \{\varnothing, \langle\omega_b\rangle, \langle\omega_s\rangle, \Omega_7\} \tag{2.4.22}$$

它是 Ω_7 上的事件 σ 域。显然 $(\Omega_7, \mathscr{F}_7)$ 是 $\mathscr{E}_7$ 产生的随机试验。

容易看出，X 的划分 Ω_5 是 Ω_7 的加细。由引理 2.4.1 得出 $(\Omega_5, \mathscr{F}_7)$ 是随机试验。显然它也是 $\mathscr{E}_7$ 产生的随机试验。

直觉上，实验 $\mathscr{E}_6$ 和 $\mathscr{E}_7$ 是 $\mathscr{E}_5$ 的一部分，不妨称它们为 $\mathscr{E}_5$ 的子实验。这个事实在自然公理系统中表现为：$(\Omega_6, \mathscr{F}_6)$，$(\Omega_7, \mathscr{F}_7)$ 是 $(\Omega_5, \mathscr{F}_5)$ 的子试验；$(\Omega_5, \mathscr{F}_6)$，$(\Omega_5, \mathscr{F}_7)$ 是 $(\Omega_5, \mathscr{F}_5)$ 的同因子试验。在 $(\Omega_5, \mathscr{F}_5)$ 中有

$$\langle\omega_o\rangle = \{1,3,5\}, \langle\omega_e\rangle = \{2,4,6\} \tag{2.4.23}$$

$$\langle\omega_b\rangle = \{5,6\}, \langle\omega_s\rangle = \{1,2,3,4\} \tag{2.4.24}$$

随机试验 $(\Omega_5, \mathscr{F}_6)$ 和 $(\Omega_5, \mathscr{F}_7)$ 证实例 1 的脚注中提到的事

实：基本事件可以不是 $\mathscr{F}$ 中的事件。

例 8 袋中有 n 个球，分别标号为 $\omega_1,\omega_2,\cdots,\omega_n$。现在随意地摸出一个球（实验 $\mathscr{E}_8$），试求 $\mathscr{E}_8$ 产生的随机试验。

解 用 ω_i 表示事件"摸出球 ω_i"。由于非空的诸事件 $\omega_i(1\leqslant i\leqslant n)$ 互不相容，其并为必然事件。故

$$\Omega_8 = \{\omega_1,\omega_2,\cdots,\omega_n\} \tag{2.4.25}$$

是原因空间 X 的一个划分。即 Ω_8 是本例的样本空间。令

$$\mathscr{F}_8 = \Omega_8 \text{ 的所有子集组成的集合} \tag{2.4.26}$$

它是 Ω_8 上的事件 σ 域。显然 $(\Omega_8,\mathscr{F}_8)$ 是 $\mathscr{E}_8$ 产生的随机试验。

例 9 接连地投掷一枚硬币，直到第一次出现正面时停止（实验 $\mathscr{E}_9$）。试求 $\mathscr{E}_9$ 产生的随机试验。

解 用 $\omega_n=(\underbrace{\text{反},\cdots,\text{反}}_{n-1\text{个}},\text{正})$ 表示事件"前 $n-1$ 次出现反面，第 n 次出现正面"，$\omega_\infty=(\text{反},\text{反},\cdots)$ 表示事件"投掷永不停止"。由于诸事件 $\omega_i(1\leqslant i\leqslant\infty)$ 非空，互不相容，其并为必然事件。故

$$\Omega_9 = \{\omega_1,\omega_2,\cdots,\omega_\infty\} \tag{2.4.27}$$

是 $\mathscr{E}_9$ 的样本空间。令

$$\mathscr{F}_9 = \Omega_9 \text{ 的所有子集组成的集合} \tag{2.4.28}$$

显然，$\mathscr{F}_9$ 是 Ω_9 上的事件 σ 域，$(\Omega_9,\mathscr{F}_9)$ 是 $\mathscr{E}_9$ 产生的随机试验。

例 10 向直线 $\mathbf{R}=(-\infty,+\infty)$ 上投掷一个质点（实验 $\mathscr{E}_{10}$）。试求 $\mathscr{E}_{10}$ 产生的随机试验。

解 用实数 x 表示非空事件"质点落在坐标为 x 的点处"。显然，诸事件互不相容（即对任意两个实数 $x\neq y$，作为事件的运算有 $x\cap y=\varnothing$），并且诸事件中有且只有一个事件出现。故这些事件组成的集合是必然事件[①]。因此

$$\mathbf{R} = \{x \mid -\infty < x < +\infty\} \tag{2.4.29}$$

---

① 不能说"诸事件的并是必然事件"。因为这些事件的个数是不可列的，事件的不可列并是无意义的。

是原因空间 X 的一个划分。即 $\mathbf{R}$ 是本例的样本空间。

用 A 表示 $\mathbf{R}$ 的任意子集，它也表示事件“质点落在 A 内”。令

$$\mathscr{F}_{10} = \{A \mid A \subset \mathbf{R}\} \tag{2.4.30}$$

显然，它是 Ω_{10} 上的事件 σ 域，也是本例所关心的全部事件。因此 $(\mathbf{R},\mathscr{F}_{10})$ 是 $\mathscr{E}_{10}$ 产生的随机试验。

例 11 向 n 维欧氏空间 $\mathbf{R}^n$ 投掷一个质点（实验 $\mathscr{E}_{11}$）。试求 $\mathscr{E}_{11}$ 产生的随机试验。

解 用 $(x_1,x_2,\cdots,x_n)$ 表示非空事件“质点落在坐标为 $(x_1,x_2,\cdots,x_n)$ 的点处”；用 $\mathbf{R}^n$ 的子集 A 表示“质点落在 A 内”。令

$$\mathbf{R}^n = \{(x_1,x_2,\cdots,x_n) \mid -\infty < x_1,x_2,\cdots,x_n < +\infty\} \tag{2.4.31}$$

$$\mathscr{F}_{11} = \{A \mid A \subset \mathbf{R}^n\} \tag{2.4.32}$$

类似于例 10 的议论得出 $(\mathbf{R}^n,\mathscr{F}_{11})$ 是 $\mathscr{E}_{11}$ 产生的随机试验。

2.4.5 随机试验建模（Ⅱ）：扩张法

设 Ω 是样本空间，$\mathscr{A}$ 是 Ω 上的一些事件组成的非空集合。

定义 2.4.5 称 $\mathscr{F}_0$ 是 $\mathscr{A}$ 生成的事件 σ 域（或事件域），如果 $\mathscr{F}_0$ 是 Ω 上包含 $\mathscr{A}$ 的事件 σ 域（或事件域），并且对 Ω 上包含 $\mathscr{A}$ 的任何事件 σ 域（或事件域）$\mathscr{F}$ 成立 $\mathscr{F} \supset \mathscr{F}_0$。

通常把事件 σ 域 $\mathscr{F}_0$ 记为 $\sigma(\mathscr{A})$，把事件域 $\mathscr{F}_0$ 记为 $r(\mathscr{A})$。$\sigma(\mathscr{A})$ 和 $r(\mathscr{A})$ 又分别称为包含 $\mathscr{A}$ 的最小事件 σ 域和最小事件域。

首先论证 $r(\mathscr{A})$ 和 $\sigma(\mathscr{A})$ 的存在性。现今概率论中流行的证明方法在自然公理系统中不能采用。该方法把证 $r(\mathscr{A})$ 和 $\sigma(\mathscr{A})$ 的存在性转变成集合论中证 $r(\mathscr{A})$ 和 $\sigma(\mathscr{A})$ 的存在性（参看定理 2.4.2），后者用到“Ω 的全体子集是 σ 域（或域）”的结论。由于这个 σ 域（或域）不一定是事件 σ 域（或域），所以不能保证 $r(\mathscr{A})$ 和 $\sigma(\mathscr{A})$ 中的集合是事件。

下面给出$r(\mathscr{A})$和$\sigma(\mathscr{A})$存在性的另一种证法[17]。它是一个构造性证明，因而清楚地显示出$r(\mathscr{A})$和$\sigma(\mathscr{A})$的结构。

定理 2.4.3 设$\mathscr{A}$是样本空间Ω上的一些事件组成的非空集合，那么$r(\mathscr{A})$和$\sigma(\mathscr{A})$皆存在且惟一。

证明 (1) $r(\mathscr{A})$的存在性和惟一性：用$\mathscr{D}$表示Ω上任意一些事件组成的集合。令

$$\mathscr{D}^* = \left\{ \bigcup_{i=1}^{m} D_i,\ \bigcap_{i=1}^{m} D_i \mid m \geqslant 1, D_i \text{ 或 } \overline{D}_i \in \mathscr{D}, 1 \leqslant i \leqslant m \right\} \tag{2.4.33}$$

即是说，$\mathscr{D}^*$是由$\mathscr{D}$首先添加$\mathscr{D}$中事件的补，然后再添加这些事件的有限并和有限交形成的集合。不妨把用$\mathscr{D}$得到$\mathscr{D}^*$的方法称为操作"$*$"。现在对$\mathscr{A}$进行一系列的"$*$"操作，令

$$\mathscr{R}_1 = \mathscr{A}, \mathscr{R}_n = \mathscr{R}_{n-1}^* \quad (n = 2,3,\cdots) \tag{2.4.34}$$

$$\mathscr{R}_0 = \bigcup_{n=1}^{\infty} \mathscr{R}_n \tag{2.4.35}$$

显然$\mathscr{R}_0$中的元素都是Ω上的事件。下证它是事件域。事实上，设$A \in \mathscr{R}_0$，由(2.4.35)得出存在自然数n使$A \in \mathscr{R}_n$。应用操作"$*$"的定义推出$\overline{A} \in \mathscr{R}_{n+1}$，得证$\overline{A} \in \mathscr{R}_0$。又设$A_1, A_2, \cdots, A_n \in \mathscr{R}_0$，则存在自然数$s_1, s_2, \cdots, s_n$使得$A_i \in \mathscr{R}_{s_i} (i = 1,2,\cdots,n)$。令$m = \max\{s_1, s_2, \cdots, s_n\}$，由操作"$*$"的定义推出$\bigcup_{i=1}^{n} A_i \in \mathscr{R}_{m+1}$。得证$\bigcup_{i=1}^{n} A_i \in \mathscr{R}_0$，故$\mathscr{R}_0$是事件域。

用$\mathscr{R}$表示包含$\mathscr{A}$的任意事件域。(2.4.34)含有$\mathscr{R}_1 \subset \mathscr{R}$。由操作"$*$"的定义知$\mathscr{R}_2, \mathscr{R}_3 \cdots$皆包含在$\mathscr{R}$内，故$\mathscr{R}_0 \subset \mathscr{R}$。得证$r(\mathscr{A})$存在且惟一，并且就是$\mathscr{R}_0$。

(2) $\sigma(\mathscr{A})$的存在性和惟一性：令

$$\mathscr{D}^{**} = \left\{ \bigcup_{i \in I} D_i,\ \bigcap_{i \in I} D_i \mid D_i \text{ 或 } \overline{D}_i \in \mathscr{D} \right\} \tag{2.4.36}$$

其中 I 是本书约定的有限或可列指标集。即是说，由 $\mathscr{D}$ 扩大为 $\mathscr{D}^*$ 时还应添加事件的可列并和可列交，得到比 $\mathscr{D}^*$ 更大的集合 $\mathscr{D}^{**}$。不妨把用 $\mathscr{D}$ 得到 $\mathscr{D}^{**}$ 的方法称为操作"* *"。现在对 $\mathscr{A}$ 进行一系列"* *"操作，令

$$\mathscr{F}_1 = \mathscr{A}, \mathscr{F}_n = \mathscr{F}_{n-1}^{**} \quad (n = 2,3,\cdots) \qquad (2.4.37)$$

用 w 表示第一个无限序数[①]，令

$$\mathscr{F}_w = \bigcup_{n=1}^{\infty} \mathscr{F}_n, \mathscr{F}_{w+1} = \mathscr{F}_w^{**}, \mathscr{F}_{w+2} = \mathscr{F}_{w+1}^{**}, \cdots\cdots \qquad (2.4.38)$$

一般情形，设 α 是任意的可数序数，并且 $\mathscr{F}_\beta(\beta<\alpha)$ 皆已定义。如果 α 是孤立序数，那么存在序数 α' 使 $\alpha'+1=\alpha$，这时规定

$$\mathscr{F}_\alpha = \mathscr{F}_{\alpha'}^{**} \qquad (2.4.39)$$

如果 α 是极限序数，则规定

$$\mathscr{F}_\alpha = \bigcup_{\beta<\alpha} \mathscr{F}_\beta \qquad (2.4.40)$$

由超限归纳法知对一切可数序数 α 皆定义了 $\mathscr{F}_\alpha$。于是得到事件集合的超限序列

$$\mathscr{F}_1, \mathscr{F}_2, \cdots, \mathscr{F}_n, \cdots, \mathscr{F}_w, \mathscr{F}_{w+1}, \cdots, \mathscr{F}_\alpha, \mathscr{F}_{\alpha+1}, \cdots \qquad (2.4.41)$$

众所周知，对任意的 α，$\{\mathscr{F}_1, \mathscr{F}_2, \cdots, \mathscr{F}_\alpha\}$ 是可列集。但是(2.4.41)的总个数是不可列的。令

$$\mathscr{F}_0 = \bigcup_\alpha \mathscr{F}_\alpha \qquad (2.4.42)$$

显然，$\mathscr{F}_0$ 中的元素是 Ω 上的事件。下证 $\mathscr{F}_0$ 是事件 σ 域。事实上，设 $A\in\mathscr{F}_0$，由(2.4.42)看出存在可数序数 α 使得 $A\in\mathscr{F}_\alpha$。由操作"* *"的定义得出 $\overline{A}\in\mathscr{F}_{\alpha+1}$，得证 $\overline{A}\in\mathscr{F}_0$。又设 $A_i\in\mathscr{F}_0$ $(i\in I)$，则存在可数序数 α_i，使得 $A_i\in\mathscr{F}_{\alpha_i}(i\in I)$。用 β 表示有限个

---

① 关于序数的知识参看[17]。在集合论中用 ω 表示第一个无限序数。由于概率论中用 ω 表示样本点，所以我们用 w 表示第一个无限序数。

或可列个序数 $\alpha_i, i\in I$ 后的第一个序数。由序数理论知 β 也是可数序数。应用操作"* *"的定义推出 $\bigcup_{i\in I} A_i \in \mathscr{F}_{\beta+1}$，得证 $\bigcup_{i\in I} A_i \in \mathscr{F}_0$。故 $\mathscr{F}_0$ 是事件 σ 域。

设 $\mathscr{F}$ 是包含 $\mathscr{A}$ 的任意事件 σ 域。由(2.4.37)看出 $\mathscr{F}_1\subset\mathscr{F}$。用超限归纳法可证，对任意的可数序数 α 有 $\mathscr{F}_\alpha\subset\mathscr{F}$。从而 $\mathscr{F}_0\subset\mathscr{F}$，得证 $\sigma(\mathscr{A})$ 存在且惟一，并且就是 $\mathscr{F}_0$。 **证毕**

推论 设 $\mathscr{A}_1$ 和 $\mathscr{A}_2$ 都是 Ω 上的一些事件组成的非空集合。如果 $\mathscr{A}_1$ 中的事件都可以用 $\mathscr{A}_2$ 中事件的补、并、交、差表达，那么

$$\sigma(\mathscr{A}_1) \subset \sigma(\mathscr{A}_2) \tag{2.4.43}$$

容易看出，用操作"* *"构造 $\sigma(\mathscr{A})$ 时可能出现情况：存在某个可数序数 α 使得 $\mathscr{F}_\alpha = \mathscr{F}_{\alpha+1} = \cdots\cdots$。例如，当 $\mathscr{A}$ 本身是事件 σ 域时有 $\mathscr{F}_1 = \mathscr{F}_2 = \cdots\cdots$。但是，存在这样的 Ω 和 $\mathscr{A}$，使得对任意的可数序数 α 成立 $\mathscr{F}_0 \neq \bigcup_{\beta\leqslant\alpha} \mathscr{F}_\beta$[17]。

定义 2.4.6 如果 $\sigma(\mathscr{A})$ 中的事件是我们关心的全部事件，则称 $\mathscr{A}$ 为芽事件集，简称为芽集。$\mathscr{A}$ 中的事件称为芽事件。

例 12 向直线 $\mathbf{R}=(-\infty,+\infty)$ 上投掷一个质点，芽集为

$$\mathscr{A}_{12} = \{(a,b] \mid -\infty \leqslant a < b < +\infty\} \tag{2.4.44}$$

其中 $(a,b]$ 表示事件"质点落在区间 $(a,b]$ 内"(实验 $\mathscr{E}_{12}$)，试求 $\mathscr{E}_{12}$ 产生的随机试验。

解 由例 10 的讨论得知 $\mathbf{R}$ 是 $\mathscr{E}_{12}$ 的样本空间。显然，芽事件皆是 $\mathbf{R}$ 上的事件。由题意，现在关心的所有事件组成的集合是 $\sigma(\mathscr{A}_{12})$。故 $(\mathbf{R},\sigma(\mathscr{A}_{12}))$ 是 $\mathscr{E}_{12}$ 产生的随机试验。

例 13 向直线 $\mathbf{R}$ 投掷一个质点。芽集为

$$\mathscr{A}_{13} = \{(a,a+1] \mid a \text{ 为整数}\} \tag{2.4.45}$$

其中 $(a,a+1]$ 表示事件"质点落在区间 $(a,a+1]$ 内"(实验 $\mathscr{E}_{13}$)。试求 $\mathscr{E}_{13}$ 产生的随机试验。

解 若选 $\mathbf{R}$ 为样本空间，显然芽事件皆是 $\mathbf{R}$ 上的事件，故 $\sigma(\mathscr{A}_{13})$ 是 $\mathbf{R}$ 上的事件 σ 域。依题意得出 $(\mathbf{R},\sigma(\mathscr{A}_{13}))$ 是 $\mathscr{E}_{13}$ 产生的随机试验。

例 14 向欧氏平面 $\mathbf{R}^2=\{(x,y)\mid x,y\in\mathbf{R}\}$ 投掷一个质点，芽集为

$$\mathscr{A}_{14}=\{(a_1,b_1]\times(a_2,b_2]\mid -\infty\leqslant a_i<b_i<+\infty, i=1,2\} \tag{2.4.46}$$

其中 $(a_1,b_1]\times(a_2,b_2]$ 表示事件"质点落在矩形 $(a_1,b_1]\times(a_2,b_2]$ 中"(实验 $\mathscr{E}_{14}$)。试求 $\mathscr{E}_{14}$ 产生的随机试验。

解 和例 10 同样的理由得出 $\mathbf{R}^2$ 是 $\mathscr{E}_{14}$ 产生的样本空间。由题意，现在关心的所有事件组成的集合是 $\sigma(\mathscr{A}_{14})$，它是 $\mathbf{R}^2$ 上的事件 σ 域。故 $(\mathbf{R}^2,\sigma(\mathscr{A}_{14}))$ 是 $\mathscr{E}_{14}$ 产生的随机试验。

例 15 向欧氏空间 $\mathbf{R}^3=\{(x,y,z)\mid x,y,z\in\mathbf{R}\}$ 投掷一个质点，芽集为

$$\mathscr{A}_{15}=\left\{\prod_{i=1}^{3}(a_i,b_i]\mid -\infty\leqslant a_i<b_i<+\infty, i=1,2,3\right\} \tag{2.4.47}$$

其中 $\prod_{i=1}^{3}(a_i,b_i]$ 表示事件"质点落在长方体 $\prod_{i=1}^{3}(a_i,b_i]$ 内"(实验 $\mathscr{E}_{15}$)。试求 $\mathscr{E}_{15}$ 产生的随机试验。

解 和例 10 同样的理由得出 $\mathbf{R}^3$ 是 $\mathscr{E}_{15}$ 产生的样本空间。由题意，$\mathscr{E}_{15}$ 关心的所有事件组成的集合是 $\sigma(\mathscr{A}_{15})$，它是 $\mathbf{R}^3$ 上的事件 σ 域。故 $(\mathbf{R}^3,\sigma(\mathscr{A}_{15}))$ 是 $\mathscr{E}_{15}$ 产生的随机试验。

2.4.6 注记

(1) 了解 $r(\mathscr{A})$ 和 $\sigma(\mathscr{A})$ 的结构对理解概率论极有帮助。它们的结构又一次显示有限和可列在概率论中处于同样的地位，不可列才是"真正的无限"。

(2) 例 10 与例 12 是两个不同的随机试验。在今后的研究中将显示出本质的差别，参看下节的定理 2.5.8 和 2.8 节的节末注记。

2.5 几类典型的随机试验

用列举法和扩张法能够构造大量的随机试验。在各种概率论的入门书籍中也能找到许许多多的例子。本节用代数学中的同构概念把常见的随机试验归并为几个典型的类。

2.5.1 σ 同构；标准试验

定义 2.5.1 设 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 是两个事件 σ 域。如果存在一个 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 之间的一一对应

$$\varphi: \quad \mathscr{F}_1 \ni A \longleftrightarrow \varphi(A) \in \mathscr{F}_2 \tag{2.5.1}$$

使得对任意的 $A, A_i \in \mathscr{F}_1 (i \in I)$ 成立

$$\varphi(\overline{A}) = \overline{\varphi(A)} \tag{2.5.2}$$

$$\varphi\Big(\bigcup_{i\in I} A_i\Big) = \bigcup_{i\in I} \varphi(A_i) \tag{2.5.3}$$

则称 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 为 σ 同构，称 φ 为 σ 同构映射。

定义 2.5.2 称随机试验 $(\Omega_1, \mathscr{F}_1)$ 和 $(\Omega_2, \mathscr{F}_2)$ 为 σ 同构，如果 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 是 σ 同构的事件 σ 域。

引理 2.5.1 设 φ 是事件 σ 域 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 之间的 σ 同构映射。那么对任意的 $A, B, A_i \in \mathscr{F}_1 (i \in I)$ 成立

(1) $\varphi(X) = X; \varphi(\varnothing) = \varnothing$ (2.5.4)

(2) $\varphi\Big(\bigcap_{i\in I} A_i\Big) = \bigcap_{i\in I} \varphi(A_i)$ (2.5.5)

(3) $\varphi(A \backslash B) = \varphi(A) \backslash \varphi(B)$ (2.5.6)

证明是容易的，留给读者。显然，2.4 节中例 1，例 3，例 4，例 6

和例 7 中的试验皆相互 σ 同构。

定义 2.5.3 称随机试验$(\Omega,\mathscr{F})$为标准的，如果对任意的 $\omega\in\Omega$ 成立$\{\omega\}\in\mathscr{F}$。

通常，用实验产生随机试验时往往得到标准试验。例如上节中的$(\Omega_i,\mathscr{F}_i)(i=1,2,\cdots,9)$，$(\mathbf{R},\mathscr{F}_{10})$，$(\mathbf{R}^n,\mathscr{F}_{11})$，$(\mathbf{R},\sigma(\mathscr{A}_{12}))$，$(\mathbf{R}^2,\sigma(\mathscr{A}_{14}))$和$(\mathbf{R}^3,\sigma(\mathscr{A}_{15}))$皆是标准试验。但是，有时也会得到非标准试验。例如上节中的$(\Omega_5,\mathscr{F}_6)$，$(\Omega_5,\mathscr{F}_7)$和$(\mathbf{R},\sigma(\mathscr{A}_{13}))$。

试问：对任意的随机试验$(\Omega,\mathscr{F})$，是否存在样本空间 Ω'，使得$(\Omega',\mathscr{F})$是标准试验？

答案是否定的。事实上，对因果空间$(X,\mathscr{U})$就没有这样的结论(参看 2.3 节中关于原因点的脚注和节末注记 3)。本节证明，常见的典型试验都是标准的，或者可化为标准的。

2.5.2 *n* 型试验

定义 2.5.4 称$(\Omega,\mathscr{F})$为 n 型试验，如果 Ω 是 n 元集。

定理 2.5.1 n 型试验$(\Omega,\mathscr{F})$是标准的当且仅当 $\mathscr{F}$ 是 Ω 的所有子集组成的集合。

证明 充分性显然。设$(\Omega,\mathscr{F})$是 n 型标准试验。不失一般性可假定

$$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\} \tag{2.5.7}$$

由标准性假定可得$\{\omega_i\}\in\mathscr{F}$。设 A 是 Ω 的任意子集，由于 $A=\bigcup\limits_{\omega\in A}\{\omega\}$ 是有限并，故 $A\in\mathscr{F}$。必要性得证。 **证毕**

定理 2.5.2 n 型标准试验皆相互 σ 同构。

证明 设$(\Omega,\mathscr{F})$和$(\Omega',\mathscr{F}')$是两个 n 型标准试验。不妨假定 Ω 有形式(2.5.7)，Ω' 有形式

$$\Omega'=\{\omega_1',\omega_2',\cdots,\omega_n'\} \tag{2.5.8}$$

任取 Ω 到 Ω' 的一个一一映射，譬如取

$$\varphi: \quad \omega_i \longleftrightarrow \omega_i' \tag{2.5.9}$$

令

$$\varphi: \quad A \in \mathscr{F} \longleftrightarrow \varphi(A) = \{\omega_i' \mid \omega_i \in A\} \in \mathscr{F}' \tag{2.5.10}$$

容易验证,(2.5.10)规定的 φ 是 $\mathscr{F}$ 和 $\mathscr{F}'$ 之间的 σ 同构映射。得证 $(\Omega,\mathscr{F})$ 和 $(\Omega',\mathscr{F}')$ 是 σ 同构的。 **证毕**

n 型标准试验 $(\Omega,\mathscr{F})$ 有清楚的结构:

(1) 样本空间 Ω 含有 n 个元素。一般形式为(2.5.7)。

(2) 事件 σ 域 $\mathscr{F}$ 含有 2^n 个元素。它们可以分为 $n+1$ 个类。第 $i(0\leqslant i\leqslant n)$ 类事件为由 i 个样本点组成的事件,这样的事件有 C_n^i 个,它们是

$$\{\omega_1,\omega_2,\cdots,\omega_i\},\{\omega_2,\omega_3,\cdots,\omega_{i+1}\},\cdots,\{\omega_{n-i+1},\omega_{n-i+2},\cdots,\omega_n\} \tag{2.5.11}$$

为了显示 n 型非标准试验的结构,需要引进

定义 2.5.5 设 $(\Omega,\mathscr{F})$ 是任意的随机试验,B 是 $\mathscr{F}$ 中的非空事件。称 B 为原子的,如果对任意的 $A\in\mathscr{F}$,由 $A\subset B$ 推出 $A=\varnothing$ 和 $A=B$ 之中必有一个成立。否则称 B 为非原子的。

显然,不同的原子事件 B_1 和 B_2 是不相容的。事实上,如果 $B_1\cap B_2\neq\varnothing$,由原子性推出 $B_1\cap B_2=B_1=B_2$,此为不可能。

定理 2.5.3 设 $(\Omega,\mathscr{F})$ 是 n 型试验。那么存在样本空间 Ω' 使得 $(\Omega',\mathscr{F})$ 是 m 型标准试验,这里 $m\leqslant n$。

证明 设 A 是 $\mathscr{F}$ 中的非空事件,则存在某原子事件 B 使得 $B\subset A$。事实上,如果 A 本身是原子的,则结论正确。否则,A 存在非空真子集 A_1 和 $A\backslash A_1$。如果它们之中有一个是原子的,把它取为 B 即得结论。否则,对 A_1 进行类似的操作。如此继续,操作停止时即得到所需要的 B。注意到 A 是有限集,所以操作在有限步内会停止。

用 $B_1,B_2,\cdots,B_m$ 表示 $\mathscr{F}$ 中所有的原子事件。令

$$\Omega' = \{B_1,B_2,\cdots,B_m\} \tag{2.5.12}$$

显然诸 $B_i(1\leqslant i\leqslant m)$非空，而且互不相交，因此 $m\leqslant n$。下证

$$\Omega=\bigcup_{i=1}^{m}B_i \tag{2.5.13}$$

事实上，若它不成立，则 $\Omega\setminus\bigcup_{i=1}^{m}B_i$ 是 $\mathscr{F}$ 中的非空事件。由已证的结论知，其内含有原子事件 B。显然 B 与诸 $B_i(1\leqslant i\leqslant m)$不相同。此为不可能，得证(2.5.13)成立。

于是得证 Ω' 是原因空间 X 的一个划分，它是 Ω 的加粗。取 Ω' 为样本空间。对任意的 $A\in\mathscr{F}$，我们有

$$A=A\cap\left[\bigcup_{i=1}^{m}B_i\right]=\bigcup_{i=1}^{m}AB_i \tag{2.5.14}$$

由 B_i 的原子性推出 $AB_i=\varnothing$ 和 $AB_i=B_i$ 之中必有一个成立。得证 A 是 Ω' 上事件，$\mathscr{F}$ 是 Ω' 上的事件 σ 域。故$(\Omega',\mathscr{F})$是 m 型标准试验。 **证毕**

2.5.3 可列型试验

定义 2.5.6 称$(\Omega,\mathscr{F})$为可列型试验，如果 Ω 含有可列个元素。

类似于 n 型试验的讨论易证

定理 2.5.4 可列型试验$(\Omega,\mathscr{F})$是标准的当且仅当 $\mathscr{F}$ 是 Ω 的所有子集组成的集合。

定理 2.5.5 可列型标准试验皆相互 σ 同构。

同样，可列型标准试验$(\Omega,\mathscr{F})$也有清楚的结构：

(1) 样本空间 Ω 含有可列个元素。不妨设

$$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n,\cdots\} \tag{2.5.15}$$

(2) 事件 σ 域 $\mathscr{F}$ 含有不可列个元素。它们可以分为可列个类：

① 第 $i(0\leqslant i<\infty)$类事件：由 i 个样本点组成的事件，它们是

$$\{\omega_1,\cdots,\omega_{i-1},\omega_i\},\{\omega_1,\cdots,\omega_{i-1},\omega_{i+1}\},\cdots \qquad (2.5.16)$$

其中第 0 类只含一个事件——不可能事件；第 $i(1\leqslant i<\infty)$ 类含有可列个事件。

② 第 w 类事件：由可列个样本点组成的事件，其个数是不可列的。

下面讨论可列型非标准试验。为此先证

引理 2.5.2 设$(\Omega,\mathscr{F})$是可列型试验。那么对任意的 $\omega\in\Omega$，存在原子事件 B 使得 $\omega\in B$。并且 B 是惟一的。

证明 如果 Ω 是原子的，取 $B=\Omega$ 即得存在性结论。否则，存在 Ω 的非空真子集 $A_1,\Omega\backslash A_1\in\mathscr{F}$。显然 $\omega\in A_1$ 和 $\omega\in\Omega\backslash A_1$ 之中有且仅有一个成立。不失一般性，设 $\omega\in A_1$。

如果 A_1 是原子事件，取 $B=A_1$ 即得存在性结论。否则，用 A_1 代替 Ω 进行上面的操作得到 A_2。如此继续，得到 $A_3,A_4,\cdots$。如果操作在自然数 k 处停止，取 $B=A_k$ 即得存在性结论。否则，应用归纳法得到事件的序列

$$\Omega,A_1,A_2,\cdots,A_n,\cdots \qquad (2.5.17)$$

令 $A_w=\bigcap_{n=1}^{\infty}A_n$。操作法保证 $\omega\in A_w$。如果 A_w 是原子的，取 $B=A_w$ 即得存在性结论。否则，应用超限归纳法继续进行操作：当 α 是孤立的可数序数时 A_α 按 A_1 的方式进行；当 α 是极限的可数序数时 A_α 按 A_w 的方式进行。于是得到事件的超限序列

$$\Omega,A_1,A_2,\cdots,A_w,\cdots,A_\alpha,\cdots \qquad (2.5.18)$$

操作法保证 $\omega\in A_\alpha,A_{\alpha+1}\backslash A_\alpha\neq\varnothing$。注意到 Ω 是可列集，而全体可数序数是不可列的，因此必存在可数序数 β，使得操作在 β 处停止。即是说，$\omega\in A_\beta,A_\beta$ 是原子事件。于是取 $B=A_\beta$ 即得存在性结论。

下证惟一性。设 B_1 和 B_2 是含 ω 的两个原子事件。由于 $B_1\cap B_2\supset\{\omega\}\neq\varnothing$，于是由 $B_1\cap B_2\subset B_1$ 和 $B_1\cap B_2\subset B_2$ 得出 $B_1\cap B_2=B_1=B_2$。

证毕

定理 2.5.6 设$(\Omega,\mathscr{F})$是可列型试验。那么存在样本空间Ω'使得$(\Omega',\mathscr{F})$是 n 型或可列型标准试验。

证明 引理 2.5.2 保证$\mathscr{F}$中存在原子事件。用$B_1,B_2,\cdots,B_n,\cdots$表示所有的原子事件。注意到诸$B_i(i\in I)$非空，并且互不相容，故其总数或为有限个(不妨设为 n 个)，或为可列个。令

$$\Omega' = \{B_1,B_2,\cdots,B_n,\cdots\} \tag{2.5.19}$$

类似于定理 2.5.3 的讨论可得，Ω'是样本空间，$(\Omega',\mathscr{F})$是标准试验。 **证毕**

容易看出，n 型试验和可列型试验具有非常类似的性质和研究方法①。因此，人们把这两类试验合称为离散型试验。

2.5.4 *n* 维波莱尔(É. Borel)试验

定义 2.5.7 (1) 称$(\mathbf{R}^n,\mathscr{B}^n)$为 n 维波莱尔试验，如果$\mathbf{R}^n$是 n 维欧氏空间，$\mathscr{B}^n=\sigma(\mathscr{A}_n)$，其中芽集为

$$\mathscr{A}_n = \left\{\prod_{i=1}^{n}(a_i,b_i] \mid -\infty\leqslant a_i<b_i<+\infty, i=1,2,\cdots,n\right\} \tag{2.5.20}$$

(2) 称$(D,\mathscr{B}^n(D))$为 D 上的波莱尔试验，如果$D\in\mathscr{B}^n$和

$$\mathscr{B}^n(D) = \{B \mid B\in\mathscr{B}^n, B\subset D\} \tag{2.5.21}$$

2.4 节中例 12、例 14 和例 15 分别是一维、二维和三维波莱尔试验。用$\langle\cdots\rangle$表示四种区间$(\cdots]$，$(\cdots)$，$[\cdots)$，$[\cdots]$中任一种。令

$$\mathscr{A}_n^* = \left\{\prod_{i=1}^{n}\langle a_i,b_i\rangle \mid -\infty\leqslant a_i<b_i<+\infty, i=1,2,\cdots,n\right\} \tag{2.5.22}$$

定理 2.5.7 $(\mathbf{R}^n,\mathscr{B}^n)$是标准试验。并且$\mathscr{B}^n=\sigma(\mathscr{A}_n^*)$。

证明 对任意的$(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$，我们有

---

① 它们之间的差别来源于处理有限和可列的差别。

$$\{(x_1,x_2,\cdots,x_n)\}=\bigcap_{m=1}^{\infty}\left(\prod_{i=1}^{n}\left(x_i-\frac{1}{m},x_i\right]\right) \quad (2.5.23)$$

应用(2.5.20)得出$\{(x_1,x_2,\cdots,x_n)\}\in\mathscr{B}^n$,得证波莱尔试验$(\mathbf{R}^n,\mathscr{B}^n)$是标准的。

容易验证

$$\prod_{i=1}^{n}(a_i,b_i)=\bigcup_{m=1}^{\infty}\left(\prod_{i=1}^{n}\left(a_i,b_i-\frac{1}{m}\right]\right) \quad (2.5.24)$$

$$\prod_{i=1}^{n}[a_i,b_i)=\bigcap_{m=1}^{\infty}\left(\prod_{i=1}^{n}\left(a_i-\frac{1}{m},b_i\right)\right) \quad (2.5.25)$$

$$\prod_{i=1}^{n}[a_i,b_i]=\bigcap_{m=1}^{\infty}\left(\prod_{i=1}^{n}\left(a_i-\frac{1}{m},b_i\right]\right) \quad (2.5.26)$$

各种混合形式的柱体(例如$(a_1,b_1)\times(a_2,b_2]\times[a_3,b_3]\times\cdots\times[a_n,b_n)$等)也有类似的等式。即是说,$\mathscr{A}_n^*$ 中的事件都可以用 $\mathscr{A}_n$ 中事件的补、并、交、差表达。应用定理 2.4.3 的推论得出 $\sigma(\mathscr{A}_n^*)\subset\sigma(\mathscr{A}_n)=\mathscr{B}^n$。另一方面,由 $\mathscr{A}_n\subset\mathscr{A}_n^*$ 推出 $\sigma(\mathscr{A}_n)\subset\sigma(\mathscr{A}_n^*)$。联合两个包含式即得 $\mathscr{B}^n=\sigma(\mathscr{A}_n^*)$。 **证毕**

这个定理表明 $\mathscr{B}^n$ 包含 $\mathbf{R}^n$ 中各种各样的柱体(或柱事件),因而也包含这些柱体用补、并、交、差产生的区域。事实上,常见的区域皆属于 $\mathscr{B}^n$,要找到一个不属于 $\mathscr{B}^n$ 的区域是相当困难的。但是,这样的区域的确存在。

定理 2.5.8 存在 $\mathbf{R}^n$ 的子集 M 使得 $M\overline{\in}\mathscr{B}^n$。

这个结论对理解波莱尔试验是非常重要的。证明用到较专门的测度论知识[16,17],从略。

2.5.5 可列维和任意维波莱尔试验

设 $\mathbf{R}$ 是实数集,$\mathscr{B}$ 是 $\mathbf{R}$ 上的波莱尔域。称

$$\mathbf{R}^{\infty}=\{(x_1,x_2,\cdots,x_i,\cdots)\mid x_i\in\mathbf{R},i=1,2,3,\cdots\} \quad (2.5.27)$$

为可列维欧氏空间。显然，对任意的 $A_i \in \mathscr{B}(i=1,2,3\cdots)$，

$$\prod_{i=1}^{\infty} A_i = \{(x_1, x_2, \cdots, x_i \cdots) | x_i \in A_i, i = 1,2,3,\cdots\} \tag{2.5.28}$$

是 $\mathbf{R}^{\infty}$ 的子集。如果诸 A_i 中除有限个(譬如说，$A_{\alpha_1}, A_{\alpha_2}, \cdots, A_{\alpha_n}$)外皆有 $A_i = \mathbf{R}(i \neq \alpha_1, \alpha_2, \cdots, \alpha_n)$，则称 $\prod_{i=1}^{\infty} A_i$ 为 n 维柱集，简记为

$$\prod_{i=1}^{n} A_{\alpha_i} \times \mathbf{R} \tag{2.5.29}$$

用 $\mathscr{A}_{\infty}$ 表示全体柱集组成的集合，即

$$\mathscr{A}_{\infty} = \left\{ \prod_{i=1}^{n} A_{\alpha_i} \times \mathbf{R} | n \geqslant 1, \alpha_i \text{ 为正整数}, \right.$$

$$\left. A_{\alpha_i} \in \mathscr{B}, i = 1,2,\cdots,n \right\} \tag{2.5.30}$$

定义 2.5.8 称 $(\mathbf{R}^{\infty}, \mathscr{B}^{\infty})$ 为可列维波莱尔试验，如果 $\mathbf{R}^{\infty}$ 是可列维欧氏空间，$\mathscr{B}^{\infty} = \sigma(\mathscr{A}_{\infty})$。

把可列维波莱尔试验推广到任意维情形意外地简单。用 T 表示任意指标集，称

$$\mathbf{R}^T = \{x(t), t \in T | \text{ 对任意固定的 } t \text{ 有 } x(t) \in \mathbf{R}\} \tag{2.5.31}$$

为 T 维欧氏空间。它是定义在 T 上的所有实值函数组成的集合。显然，对任意的 $A_t \in \mathscr{B}(t \in T)$，

$$\prod_{t \in T} A_t = \{x(t), t \in T | \text{ 对固定的 } t \in T \text{ 有 } x(t) \in A_t\} \tag{2.5.32}$$

是 $\mathbf{R}^T$ 的子集。如果诸 A_t 中除有限个(譬如说，$A_{t_1}, A_{t_2}, \cdots, A_{t_n}$)外皆有 $A_t = \mathbf{R}(t \neq t_1, t_2, \cdots, t_n)$，则称 $\prod_{t \in T} A_t$ 为 n 维柱集，并简记为

$$\prod_{i=1}^{n} A_{t_i} \times \mathbf{R} \tag{2.5.33}$$

用 $\mathscr{A}_T$ 表示全体柱集组成的集合，即

$$\mathscr{A}_T = \left\{ \prod_{i=1}^{n} A_{t_i} \times \mathbf{R} \,\middle|\, n \geqslant 1, t_i \in T, A_{t_i} \in \mathscr{B}, i = 1,2,\cdots,n \right\} \tag{2.5.34}$$

定义 2.5.9 称 $(\mathbf{R}^T, \mathscr{B}^T)$ 为 T 维波莱尔试验，如果 $\mathbf{R}^T$ 是 T 维欧氏空间，$\mathscr{B}^T = \sigma(\mathscr{A}_T)$。

2.5.6 注记

试验 $(\Omega, \mathscr{F})$ 的标准性具有重要的意义。在 K 公理系统中曾提出过类似的问题：是否存在与 $(\Omega, \mathscr{F}, P)$ σ 同构的概率空间 $(\Omega', \mathscr{F}', P')$，使得当 $\omega' \in \Omega'$ 时成立 $\{\omega'\} \in \mathscr{F}'$？

对于离散型概率空间这个问题有肯定的回答[18]。对于一般的概率空间，当把 σ 同构的要求降低为本质同构后也能得到肯定的回答[19]。我们将在 2.7 节的节末注记中给出概率空间标准化的另一种方法。

2.6 联合随机试验

在因果空间中存在无穷无尽的随机试验。这些试验不是孤立的，它们之间存在或松或紧的联系。联合随机试验是研究这类联系的一种重要的工具。

2.6.1 联合事件 σ 域

引理 2.6.1 设 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 是两个事件 σ 域。令

$$\mathscr{F}_1 * \mathscr{F}_2 = \{A \cap B \mid A \in \mathscr{F}_1, B \in \mathscr{F}_2\} \tag{2.6.1}$$

那么成立

(1) $\mathscr{F}_i \subset \mathscr{F}_1 * \mathscr{F}_2 (i=1,2)$；

(2) $\mathscr{F}_1 * \mathscr{F}_2$ 是 π 系，即如果 $C, D \in \mathscr{F}_1 * \mathscr{F}_2$，则 $C \cap D \in$

$\mathscr{F}_1 * \mathscr{F}_2$；

(3) $\sigma(\mathscr{F}_1 * \mathscr{F}_2)=\sigma(\mathscr{F}_1 \cup \mathscr{F}_2)$。

证明不难，留给读者。今后把 $\sigma(\mathscr{F}_1 * \mathscr{F}_2)$简记为 $\mathscr{F}_1 \otimes \mathscr{F}_2$。引理表明 $\mathscr{F}_1 \otimes \mathscr{F}_2$ 是包含 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 的最小事件 σ 域。

定义 2.6.1 称 $\mathscr{F}_1 \otimes \mathscr{F}_2$ 为 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 的联合事件 σ 域。称 $\mathscr{F}_1 * \mathscr{F}_2$ 为柱事件集，其内的事件 $A \cap B$ 称为柱事件，且称 A 和 B 是它的边。

联合事件 σ 域建立因果空间的一种直观形象：因果空间中的事件都存在联系，第 I 组公理用补、并、交、差表达"联系"。因此，$\mathscr{F}_1$ 和 $\mathscr{F}_2$ 之间的联系表现为 $\mathscr{F}_1$ 中事件和 $\mathscr{F}_2$ 中事件的补、并、交、差。由于 $\mathscr{F}_1 \otimes \mathscr{F}_2$ 是包含这些补事件、并事件、交事件和差事件的最小事件 σ 域，于是 $\mathscr{F}_1 \otimes \mathscr{F}_2$ 成为研究 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 之间联系的最方便场所。

2.6.2 联合随机试验

本小段论证研究随机试验$(\Omega_1, \mathscr{F}_1)$和$(\Omega_2, \mathscr{F}_2)$之间关系的最方便场所是它们的联合随机试验。

定理 2.6.1 设$(\Omega_1, \mathscr{F}_1)$和$(\Omega_2, \mathscr{F}_2)$是两个随机试验。那么$(\Omega_1 \circ \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2)$是随机试验，$(\Omega_i, \mathscr{F}_i)$ $(i=1,2)$是它的子试验。其中 $\Omega_1 \circ \Omega_2$ 是 Ω_1 和 Ω_2 的交叉划分，$\mathscr{F}_1 \otimes \mathscr{F}_2$ 是 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 的联合事件 σ 域。

证明 由于引理 2.6.1，只须证$(\Omega_1 \circ \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2)$是随机试验。不失一般性，设

$$\Omega_1 = \{\omega_\alpha^{(1)} \mid \alpha \in J_1\}, \Omega_2 = \{\omega_\beta^{(2)} \mid \beta \in J_2\} \qquad (2.6.2)$$

其中 J_1 和 J_2 是两个指标集。由定义 2.3.5 知

$$\Omega_1 \circ \Omega_2 = \{\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \mid \alpha \in J_1, \beta \in J_2, \omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \neq \varnothing\} \qquad (2.6.3)$$

任取 $A \in \mathscr{F}_1, B \in \mathscr{F}_2$。由于 A 和 B 分别是 Ω_1 和 Ω_2 上的事

件，不失一般性可设

$$A=\{\omega_\alpha^{(1)}|\alpha\in J_{1A}\};B=\{\omega_\beta^{(2)}|\beta\in J_{2B}\} \tag{2.6.4}$$

其中 J_{1A}和 J_{2B}分别是 J_1 和 J_2 的子集。在因果空间中两事件的交为

$$A\cap B=\{\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}|\alpha\in J_{1A},\beta\in J_{2B},\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}\neq\varnothing\} \tag{2.6.5}$$

由此得出它是 $\Omega_1\circ\Omega_2$ 上的事件。从而 $\sigma(\mathscr{F}_1*\mathscr{F}_2)$是 $\Omega_1\circ\Omega_2$ 上的事件 σ 域。得证$(\Omega_1\circ\Omega_2,\mathscr{F}_1\otimes\mathscr{F}_2)$是随机试验。 **证毕**

定义 2.6.2 称$(\Omega_1\circ\Omega_2,\mathscr{F}_1\otimes\mathscr{F}_2)$为$(\Omega_1,\mathscr{F}_1)$和$(\Omega_2,\mathscr{F}_2)$的联合随机试验，简称为联合试验。

显然，当$(\Omega_2,\mathscr{F}_2)$是$(\Omega_1,\mathscr{F}_1)$的子试验时，它们的联合试验是$(\Omega_1\circ\Omega_2,\mathscr{F}_1)$；如果是同因子试验，则联合试验是$(\Omega_1,\mathscr{F}_1)$。

例 1 试求 2.4 节中$(\Omega_1,\mathscr{F}_1)$和$(\Omega_3,\mathscr{F}_3)$的联合随机试验。

解 所求的联合试验为$(\Omega_1\circ\Omega_3,\mathscr{F}_1\otimes\mathscr{F}_3)$。其中

$$\Omega_1\circ\Omega_3=\{正\cap男,正\cap女,反\cap男,反\cap女\} \tag{2.6.6}$$

$$\mathscr{F}_1\otimes\mathscr{F}_3=\Omega_1\circ\Omega_3\text{ 的全体子集组成的集合} \tag{2.6.7}$$

在现实中投硬币和生小孩是毫无关联的两个实验。但是，它们的联合试验仍然具有现实意义。例如，在日常生活中有时会遇到这样的场面：某父亲用投硬币的方式猜测未来孩子性别时说"我投一枚硬币。如果出现正面，猜男孩；如果出现反面，猜女孩"。试问：如何描述父亲的行为？自然公理系统采用的方法是：它认为父亲做的随机试验是例 1 中的联合试验，但是父亲真正关心的是子试验$(\Omega_1\circ\Omega_3,\mathscr{F}^*)$，其中 $\mathscr{F}^*=\{\varnothing,A,\bar{A},\Omega_1\circ\Omega_3\}$，事件

$$A=\{正\cap男,反\cap女\};$$

$$\bar{A}=\{正\cap女,反\cap男\} \tag{2.6.8}$$

分别表示事件"父亲猜对了"和"父亲猜错了"。

例 2 验证 2.4 节中$(\Omega_5,\mathscr{F}_5)$和$(\Omega_6,\mathscr{F}_6)$的联合试验是$(\Omega_5,\mathscr{F}_5)$。

解 应用(2.4.23)得出 $\Omega_5 * \Omega_6$ 的元素为

$$\begin{cases} i \cap \omega_o = i, i \cap \omega_e = \varnothing, & 当\ i = 1,3,5 \\ i \cap \omega_o = \varnothing, i \cap \omega_e = i, & 当\ i = 2,4,6 \end{cases} \tag{2.6.9}$$

删除空集后得到 $\Omega_5 \circ \Omega_6 = \Omega_5$。显然 $\mathscr{F}_5 \otimes \mathscr{F}_6 = \mathscr{F}_5$。故所求的联合试验是$(\Omega_5, \mathscr{F}_5)$。

例 3 试求 2.4 节中$(\Omega_6, \mathscr{F}_6)$和$(\Omega_7, \mathscr{F}_7)$的联合试验。

解 所求的联合试验为$(\Omega_6 \circ \Omega_7, \mathscr{F}_6 \otimes \mathscr{F}_7)$。其中

$$\Omega_6 \circ \Omega_7 = \{\omega_o \cap \omega_b, \omega_o \cap \omega_s, \omega_e \cap \omega_b, \omega_e \cap \omega_s\} \tag{2.6.10}$$

$$\mathscr{F}_6 \otimes \mathscr{F}_7 = \Omega_6 \circ \Omega_7 \text{ 的全体子集组成的集合} \tag{2.6.11}$$

注 如果采用$(\Omega_5, \mathscr{F}_6)$和$(\Omega_5, \mathscr{F}_7)$作为 2.4 节中例 6 和例 7 产生的随机试验。那么它们的联合试验为$(\Omega_5, \mathscr{F}_6 \otimes \mathscr{F}_7)$。在这个试验中有

$$\omega_o \cap \omega_b = \{5\}, \omega_o \cap \omega_s = \{1,3\},$$
$$\omega_e \cap \omega_b = \{6\}, \omega_e \cap \omega_s = \{2,4\} \tag{2.6.12}$$

例 4 把 2.4 节例 1 的实验 $\mathscr{E}_1$ 进行两次,得到新实验 $\mathscr{E}^*$。试求 $\mathscr{E}^*$ 产生的随机试验。

解 第 1 次掷硬币和第 2 次掷硬币是不同的两个实验。产生的事件也是不同的。像"出现正面"这样的词在 $\mathscr{E}^*$ 中是含糊的,不能用它描述一个事件。事实上,它至少存在 4 种理解:① 第 1 次投掷时出现正面,可记为正$_1$;② 第 2 次投掷时出现正面,可记为正$_2$;③ 至少出现一次正面,可记为正$_1 \cup$正$_2$;④ 出现两次正面,可记为正$_1 \cap$正$_2$。

为了消除这类含糊性,把 2.4 节例 1 中产生的符号都添加下标 $i(i=1,2)$。下标 i 表示它是第 i 次投掷时的产物。例如,第 i 次投掷是实验 $\mathscr{E}_{1i}$,它产生的随机试验是$(\Omega_{1i}, \mathscr{F}_{1i})$,其中

$$\Omega_{1i} = \{正_i, 反_i\} \tag{2.6.13}$$

$$\mathscr{F}_{1i} = \{\varnothing, \langle 正_i \rangle, \langle 反_i \rangle, \Omega_{1i}\} \tag{2.6.14}$$

应用新符号,我们能够准确地表达出 $\mathscr{E}^*$ 产生的随机试验,它是

$(\Omega_{11},\mathscr{F}_{11})$和$(\Omega_{12},\mathscr{F}_{12})$的联合试验$(\Omega_{11}\circ\Omega_{12},\mathscr{F}_{11}\otimes\mathscr{F}_{12})$，其中

$$\Omega_{11}\circ\Omega_{12}=\{正_1\cap 正_2,正_1\cap 反_2,反_1\cap 正_2,反_1\cap 反_2\} \tag{2.6.15}$$

$$\mathscr{F}_{11}\otimes\mathscr{F}_{12}=\Omega_{11}\circ\Omega_{12}\text{的全体子集组成的集合} \tag{2.6.16}$$

2.6.3 乘积随机试验

乘积随机试验是联合试验的一种重要的特殊情形。它在概率论中已被广泛使用。

定义 2.6.3 如果 $\Omega_1\circ\Omega_2$ 是真交叉划分，则称联合试验$(\Omega_1\circ\Omega_2,\mathscr{F}_1\otimes\mathscr{F}_2)$为乘积随机试验，简称为乘积试验。

引理 2.6.2 设$(\Omega_1\circ\Omega_2,\mathscr{F}_1\otimes\mathscr{F}_2)$是$(\Omega_1,\mathscr{F}_1)$和$(\Omega_2,\mathscr{F}_2)$的乘积试验。那么

(1) 假定 Ω_1(或 Ω_2)不是最粗划分，则对任意的 $\omega_\alpha^{(1)}\in\Omega_1$，$\omega_\beta^{(2)}\in\Omega_2$，集合 $\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}$ 是 $\omega_\beta^{(2)}$(或 $\omega_\alpha^{(1)}$)的真子集；

(2) 柱事件 $A\cap B=\varnothing$ 当且仅当 $A=\varnothing$ 和 $B=\varnothing$ 之中至少有一个成立；

(3) 非空柱事件 $A\cap B\subset C\cap D$ 当且仅当 $A\subset C$ 且 $B\subset D$；

(4) 非空柱事件 $A\cap B=C\cap D$ 当且仅当 $A=C$ 且 $B=D$；

(5) $\mathscr{F}_1\cap\mathscr{F}_2=\{\varnothing,X\}$；

(6) 如果 $A\in\mathscr{F}_1,B\in\mathscr{F}_2$，并且 A 和 B 皆不是$\varnothing$或 X，那么柱事件

$$A\cap B\,\overline{\in}\,\mathscr{F}_1;A\cap B\,\overline{\in}\,\mathscr{F}_2 \tag{2.6.17}$$

证明 不失一般性，设 Ω_1 和 Ω_2 有表达式(2.6.2)。那么引理的条件蕴含

$$\Omega_1\circ\Omega_2=\{\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}\mid\alpha\in J_1,\beta\in J_2\} \tag{2.6.18}$$

(1) 由于 Ω_1 不是最粗划分，故 $J_1\setminus\{\alpha\}$非空。由(2.6.18)推出 $\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}$ 和$\{\omega_r^{(1)}\cap\omega_\beta^{(2)}\mid r\in J_1\setminus\{\alpha\}\}$是原因空间的两个不相交的非空子集。显然它们都是 $\omega_\beta^{(2)}$ 的子集，得证 $\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}$ 是 $\omega_\beta^{(2)}$ 的真

子集。

(2) 如果 $A\neq\varnothing$ 且 $B\neq\varnothing$，那么 Ω_1 中存在样本点 $\omega_\alpha^{(1)}\in A$，Ω_2 中存在样本点 $\omega_\beta^{(2)}\in B$。由此推出 $\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}\in A\cap B$。由 $\Omega_1\circ\Omega_2$ 是真交叉划分的假设得出柱事件 $A\cap B$ 非空，必要性得证。充分性显然。

(3) 假定 $A\cap B\subset C\cap D$ 成立。如果 $A\subset C$ 和 $B\subset D$ 中至少有一个不成立，不妨设 $A\subset C$ 不成立。由 $A\cap B$ 非空的假定和(2)得出 $A\neq\varnothing$。因此，Ω_1 中存在样本点 $\omega_\alpha^{(1)}$ 使得 $\omega_\alpha^{(1)}\in A$ 且 $\omega_\alpha^{(1)}\overline{\in}C$。

同理，$B\neq\varnothing$。故 Ω_2 中存在样本点 $\omega_\beta^{(2)}\in B$。由此推出

$$\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}\in A\cap B \text{ 且 } \omega_\alpha^{(1)}\cap\omega_\beta^{(2)}\overline{\in}C\cap D$$

注意到 $\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}$ 的非空性假定，由此得出 $A\cap B\subset C\cap D$ 不成立。与假定矛盾，必要性得证。充分性显然。

(4) 结论由(3)直接推出。

(5) 容易验证 $\mathscr{F}_1\cap\mathscr{F}_2$ 是事件 σ 域。由定理 2.2.5 推出 $\{\varnothing,X\}\subset\mathscr{F}_1\cap\mathscr{F}_2$。下证 $\mathscr{F}_1\cap\mathscr{F}_2=\{\varnothing,X\}$。

假定 $\mathscr{F}_1\cap\mathscr{F}_2\neq\{\varnothing,X\}$。那么存在事件 A 使得

$$A\in\mathscr{F}_1\cap\mathscr{F}_2,A\neq\varnothing,A\neq X$$

由此推出

$$\overline{A}\in\mathscr{F}_1\cap\mathscr{F}_2,\overline{A}\neq\varnothing,\overline{A}\neq X$$

现在，由 $A\in\mathscr{F}_1$ 推出存在 $\omega_\alpha^{(1)}\in\Omega_1$ 使得 $\omega_\alpha^{(1)}\in A$；由 $\overline{A}\in\mathscr{F}_2$ 推出存在 $\omega_\beta^{(2)}\in\Omega_2$ 使得 $\omega_\beta^{(2)}\in\overline{A}$。故

$$\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}\in A\cap\overline{A}=\varnothing$$

与 $\Omega_1\circ\Omega_2$ 是真交叉划分矛盾，结论得证。

(6) 不失一般性，设 A 和 B 有表达式(2.6.4)。由于 A 和 B 皆不是$\varnothing$和 X，故 J_{1A}和 J_{2B}分别是 J_1 和 J_2 的非空真子集。

设 $\omega_\alpha^{(1)}\in A$ 是 Ω_1 的样本点。由(1)得出原因点的集合

$$\{\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}\mid\beta\in J_{2B}\},\{\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}\mid\beta\in J_2\backslash J_{2B}\}$$

是 $\omega_\alpha^{(1)}$ 的两个非空真子集。显然前者属于 $A\cap B$，后者不属于 $A\cap B$。因此 $A\cap B$ 不是 Ω_1 上的事件，得证 $A\cap B\overline{\in}\mathscr{F}_1$。同理可证 $A\cap B\overline{\in}\mathscr{F}_2$。 **证毕**

回忆测度论中乘积可测空间的知识[16]。设(Y,S)和(Z,T)是两个可测空间。称$(y,z)(y\in Y,z\in Z)$为序偶。全体序偶组成的集合记为$Y\times Z$，即

$$Y\times Z=\{(y,z)|y\in Y,z\in Z\} \tag{2.6.19}$$

称$Y\times Z$为Y和Z的笛卡儿(R. Descartes)乘积空间。

设$A\in S,B\in T$。称

$$A\times B=\{(y,z)|y\in A,z\in B\} \tag{2.6.20}$$

为可测矩形，且称A和B是它的边。用$\mathscr{C}$表示全体可测矩形组成的集合，即

$$\mathscr{C}=\{A\times B|A\in S,B\in T\} \tag{2.6.21}$$

称$\mathscr{C}$为S和T产生的可测矩形集。用$S\times T$表示包含$\mathscr{C}$的最小σ域，即

$$S\times T=\sigma(\mathscr{C}) \tag{2.6.22}$$

称$S\times T$为S和T的乘积σ域。

称二元组$(Y\times Z,S\times T)$为(Y,S)和(Z,T)的乘积可测空间。

定理 2.6.2 设$(\Omega_1,\mathscr{F}_1)$和$(\Omega_2,\mathscr{F}_2)$是两个随机试验，$\Omega_1\circ\Omega_2$是真交叉划分。如果用$(\omega_\alpha,\omega_\beta)$表示$\Omega_1\circ\Omega_2$中的样本点$\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}$，那么乘积试验$(\Omega_1\circ\Omega_2,\mathscr{F}_1\otimes\mathscr{F}_2)$成为$(\Omega_1\times\Omega_2,\mathscr{F}_1\times\mathscr{F}_2)$。

证明 用φ表示把$\Omega_1\circ\Omega_2$中样本点$\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}$改记为$(\omega_\alpha,\omega_\beta)$的操作。即

$$\varphi(\omega_\alpha^{(1)}\cap\omega_\beta^{(2)})=(\omega_\alpha,\omega_\beta) \tag{2.6.23}$$

比较(2.6.18)和(2.6.19)得出，φ是$\Omega_1\circ\Omega_2$和$\Omega_1\times\Omega_2$之间的1-1映射。对任意的$D\in\mathscr{F}_1\otimes\mathscr{F}_2$，它在映射$\varphi$下的象为

$$\varphi(D)=\{(\omega_\alpha,\omega_\beta)\mid\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}\in D\}\tag{2.6.24}$$

如果能证明 φ 是 $\mathscr{F}_1\otimes\mathscr{F}_2$ 和 $\mathscr{F}_1\times\mathscr{F}_2$ 之间的 σ 同构映射，则定理结论得证。

比较(2.6.1)和(2.6.21)得出，φ 是 $\mathscr{F}_1*\mathscr{F}_2$ 和 $\mathscr{C}$ 之间的1-1对应。并且满足

(1) 对任意的 $A_i\cap B_i\in\mathscr{F}_1*\mathscr{F}_2(i\in I)$ 成立

$$\varphi\Big[\bigcup_{i\in I}(A_i\cap B_i)\Big]=\bigcup_{i\in I}(A_i\times B_i)\tag{2.6.25}$$

(2) 对任意的 $A\cap B\in\mathscr{F}_1*\mathscr{F}_2$ 成立

$$\varphi(\overline{A\cap B})=\overline{A\times B}\tag{2.6.26}$$

事实上，(2.6.25)显然。应用(2.6.25)，我们有

$$\begin{aligned}\varphi(\overline{A\cap B})&=\varphi[(A\cap\overline{B})\cup(\overline{A}\cap B)\cup(\overline{A}\cap\overline{B})]\\&=(A\times\overline{B})\cup(\overline{A}\times B)\cup(\overline{A}\times\overline{B})=\overline{A\times B}\end{aligned}$$

得证(2.6.26)。

现在应用(1)和(2)容易推出，$\mathscr{F}_1*\mathscr{F}_2$ 中事件的补、并、交、差在映射 φ 下的象是相应的可测矩形的补、并、交、差。由此得出(2.6.24)规定的 φ 是 $(\mathscr{F}_1*\mathscr{F}_2)^{**}$ 和 $\mathscr{C}^{**}$ 之间的 1-1 映射，并且满足(1)和(2)，这里右上标“* *”是定理 2.4.3 中的操作。注意到 $\mathscr{F}_1\otimes\mathscr{F}_2$ 和 $\mathscr{F}_1\times\mathscr{F}_2$ 都是用操作“* *”构造的，应用超限归纳法即得 φ 是 $\mathscr{F}_1\otimes\mathscr{F}_2$ 和 $\mathscr{F}_1\times\mathscr{F}_2$ 之间的 σ 同构映射。　　**证毕**

今后用$(\Omega_1\times\Omega_2,\mathscr{F}_1\times\mathscr{F}_2)$表示$(\Omega_1,\mathscr{F}_1)$和$(\Omega_2,\mathscr{F}_2)$的乘积试验。这种表示法将带来书写上的方便。例如，用它给出例 4 的解答时可以简单地说，实验 $\mathscr{E}^*$ 产生的随机试验是$(\Omega_1\times\Omega_1,\mathscr{F}_1\times\mathscr{F}_1)$，免去添加下标“$i$”的麻烦。注意，原先的解答必须添加下标。否则产生出$(\Omega_1\circ\Omega_1,\mathscr{F}_1\otimes\mathscr{F}_1)$，按运算“$\circ$”与“$\otimes$”的规则它是$(\Omega_1,\mathscr{F}_1)$，因而不是问题的解答。

2.6.4 重复随机试验

定义 2.6.4 称$(\Omega\times\Omega,\mathscr{F}\times\mathscr{F})$为$(\Omega,\mathscr{F})$的重复随机试验。

简称为重复试验，简记为$(\Omega^2,\mathscr{F}^2)$。

定理 2.6.3 设实验$\mathscr{E}$产生标准试验$(\Omega,\mathscr{F})$。如果把实验$\mathscr{E}$重复进行两次，得到联合实验$\mathscr{E}^*$。那么$\mathscr{E}^*$产生的试验是$(\Omega^2,\mathscr{F}^2)$。

证明 不失一般性，设

$$\Omega=\{\omega_\alpha \mid \alpha\in J\} \tag{2.6.27}$$

其中J为指标集。现在把实验$\mathscr{E}$的所有产物添加下标i，表示它是第i次进行实验时的产物。特别，$(\Omega_i,\mathscr{F}_i)$是第i次实验产生的试验$(i=1,2)$。因此，联合实验$\mathscr{E}^*$产生试验$(\Omega_1\circ\Omega_2,\mathscr{F}_1\otimes\mathscr{F}_2)$。

应用定理 2.6.2，为完成定理的证明，只须证$\Omega_1\circ\Omega_2$是真交叉划分。显然

$$\Omega_1\circ\Omega_2=\{\omega_{\alpha1}\cap\omega_{\beta2} \mid \alpha,\beta\in J,\omega_{\alpha1}\cap\omega_{\beta2}\neq\varnothing\} \tag{2.6.28}$$

由标准性假定得出$\{\omega_{\alpha1}\}\in\mathscr{F}_1$，$\{\omega_{\beta2}\}\in\mathscr{F}_2$。于是$\omega_{\alpha1}\cap\omega_{\beta2}$表示事件：第 1 次进行实验$\mathscr{E}$时出现基本事件$\omega_{\alpha1}$，第 2 次出现基本事件$\omega_{\beta2}$。这是一个可能出现的事件，并非空事件。得证$\Omega_1\circ\Omega_2$是真交叉划分。 **证毕**

例 5 二维波莱尔试验$(R^2,\mathscr{B}^2)$是一维波莱尔试验$(R,\mathscr{B})$的重复试验。

2.6.5 *n* 维联合试验

容易把二维联合试验推广到n维的情形。假定$(\Omega_i,\mathscr{F}_i)(i=1,2,\cdots,n)$是$n$个随机试验。不失一般性，设

$$\Omega_i=\{\omega_{\alpha_i}^{(i)} \mid \alpha_i\in J_i\} \tag{2.6.29}$$

其中J_i是指标集。容易证明

$$\Omega_1*\Omega_2*\cdots*\Omega_n=\{\omega_{\alpha_1}^{(i)}\cap\omega_{\alpha_2}^{(2)}\cap\cdots\cap\omega_{\alpha_n}^{(n)} \mid \alpha_i\in J_i,\ i=1,2,\cdots n\} \tag{2.6.30}$$

是原因空间 X 的一个广义划分。把其中的空集删除后记为 $\Omega_1 \circ \Omega_2 \circ \cdots \circ \Omega_n$，称之为 $\Omega_i(i=1,2,\cdots,n)$的交叉划分。特别，当 $\Omega_1 * \Omega_2 * \cdots * \Omega_n$ 不含空集时称 $\Omega_1 \circ \Omega_2 \circ \cdots \circ \Omega_n$ 为真交叉划分。

任取 $A_i \in \mathscr{F}_i(i=1,2,\cdots,n)$。称事件

$$A_1 \cap A_2 \cap \cdots \cap A_n = \{\omega_{\alpha_1}^{(1)} \cap \omega_{\alpha_2}^{(2)} \cap \cdots \cap \omega_{\alpha_n}^{(n)} \mid \omega_{\alpha_i}^{(i)} \in A_i, \quad i = 1,2,\cdots,n\} \tag{2.6.31}$$

为 n 维柱事件，且称 $A_i(1 \leqslant i \leqslant n)$是它的边。用 $\mathscr{F}_1 * \mathscr{F}_2 * \cdots * \mathscr{F}_n$ 表示全体柱事件组成的集合，即

$$\mathscr{F}_1 * \mathscr{F}_2 * \cdots * \mathscr{F}_n = \{A_1 \cap A_2 \cap \cdots \cap A_n \mid A_i \in \mathscr{F}_i, i = 1,2,\cdots,n\} \tag{2.6.32}$$

称为 $\mathscr{F}_1,\mathscr{F}_2,\cdots,\mathscr{F}_n$ 产生的柱事件集。用 $\mathscr{F}_1 \otimes \mathscr{F}_2 \otimes \cdots \otimes \mathscr{F}_n$ 表示包含所有柱事件的最小事件 σ 域，即

$$\mathscr{F}_1 \otimes \mathscr{F}_2 \otimes \cdots \otimes \mathscr{F}_n = \sigma(\mathscr{F}_1 * \mathscr{F}_2 * \cdots * \mathscr{F}_n) \tag{2.6.33}$$

显然，$\mathscr{F}_1 \otimes \mathscr{F}_2 \otimes \cdots \otimes \mathscr{F}_n$ 是样本空间 $\Omega_1 \circ \Omega_2 \circ \cdots \circ \Omega_n$ 上的事件 σ 域，也是包含诸 $\mathscr{F}_i(i=1,2,\cdots,n)$的最小事件 σ 域，称之为 $\mathscr{F}_1,\mathscr{F}_2,\cdots,\mathscr{F}_n$ 的联合事件 σ 域。

不难证明，二维情形的讨论适用于 n 维情形。

定义 2.6.5 称$(\Omega_1 \circ \Omega_2 \circ \cdots \circ \Omega_n, \mathscr{F}_1 \otimes \mathscr{F}_2 \otimes \cdots \otimes \mathscr{F}_n)$为诸试验$(\Omega_i, \mathscr{F}_i)(i=1,2,\cdots,n)$的 n 维联合随机试验，简称为联合试验。

两种重要的特殊情形是：

(1) 当 $\Omega_1 \circ \Omega_2 \circ \cdots \circ \Omega_n$ 是真交叉划分时称联合试验为乘积试验。这时定理 2.6.2 可以推广到 n 维情形。因此，我们用测度论中的 n 维乘积可测空间$(\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n, \mathscr{F}_1 \times \mathscr{F}_2 \times \cdots \times \mathscr{F}_n)$表示乘积试验。

(2) 称$(\underbrace{\Omega \times \Omega \times \cdots \times \Omega}_{n个}, \underbrace{\mathscr{F} \times \mathscr{F} \times \cdots \times \mathscr{F}}_{n个})$为$(\Omega, \mathscr{F})$的 n 重

试验，简记为$(\Omega^n,\mathscr{F}^n)$。这时定理6.2.3可以推广到n重情形。

例6 n维波莱尔试验$(R^n,\mathscr{B}^n)$是一维波莱尔试验$(R,\mathscr{B})$的n重试验。

2.6.6 可列维和任意维联合试验

用T表示指标集，它可以是可列集，也可以是不可列集。假定$(\Omega_t,\mathscr{F}_t)(t\in T)$是一族随机试验。不失一般性，设

$$\Omega_t = \{\omega_{\alpha_t}^{(t)} \mid \alpha_t \in J_t\} \tag{2.6.34}$$

其中J_t是指标集。容易证明

$$\mathop{*}_{t\in T}\Omega_t = \left\{ \bigcap_{t\in T} \omega_{\alpha_t}^{(t)} \mid \alpha_t \in J_t \right\} \tag{2.6.35}$$

是原因空间X的一个广义划分。把其中的空集删除后记为$\mathop{\circ}_{t\in T}\Omega_t$，称之为诸$\Omega_t(t\in T)$的交叉划分。特别，$\mathop{*}_{t\in T}\Omega_t$不含空集时称$\mathop{\circ}_{t\in T}\Omega_t$为真交叉划分。

任取$A_t\in\mathscr{F}_t(t\in T)$。如果除有限个(譬如说，$A_{t_1},A_{t_2},\cdots,A_{t_n}$)外皆有$A_t=\Omega_t(t\neq t_1,t_2,\cdots,t_n)$，则称事件

$$\bigcap_{t\in T} A_t \overset{\text{def}}{=\!=} \bigcap_{i=1}^{n} A_{t_i} = \left\{ \bigcap_{t\in T} \omega_{\alpha_t}^{(t)} \mid \omega_{\alpha_t}^{(t)} \in A_t, t\in T \right\} \tag{2.6.36}$$

为n维柱事件，称$A_{t_i}(i=1,2,\cdots,n)$为它的边。并简记为

$$\bigcap_{i=1}^{n} A_{t_i} \cap \Omega_t \tag{2.6.37}$$

显然它是样本空间$\mathop{\circ}_{t\in T}\Omega_t$上的事件。令

$$\mathop{*}_{t\in T}\mathscr{F}_t = \left\{ \bigcap_{i=1}^{n} A_{t_i} \cap \Omega_t \mid n\geqslant 1, A_{t_i}\in\mathscr{F}_{t_i}, i=1,2,\cdots,n \right\} \tag{2.6.38}$$

称之为诸$\mathscr{F}_t(t\in T)$产生的柱事件集。令

$$\bigotimes_{t\in T}\mathscr{F}_t = \sigma\left(\mathop{*}_{t\in T}\mathscr{F}_t\right) \tag{2.6.39}$$

显然，它是 $\underset{t\in T}{\circ}\Omega_t$ 上的事件 σ 域，也是包含诸 $\mathscr{F}_t(t\in T)$ 的最小事件 σ 域。称它为诸 $\mathscr{F}_t(t\in T)$ 的联合事件 σ 域。

同样，二维情形的讨论可以推广到 T 维情形。

定义 2: 6. 6 称 $(\underset{t\in T}{\circ}\Omega_t, \underset{t\in T}{\otimes}\mathscr{F}_t)$ 为诸试验 $(\Omega_t,\mathscr{F}_t)(t\in T)$ 的 T 维联合随机试验，简称为联合试验。

两种重要的特殊情形是：

(1) 当 $\underset{t\in T}{\circ}\Omega_t$ 是真交叉划分时称联合试验为乘积试验。这时定理 2. 6. 2 可以推广到 T 维情形。因此，我们用测度论中 T 维乘积可测空间 $\left(\prod_{t\in T}\Omega_t, \prod_{t\in T}\mathscr{F}_t\right)$ 表示 T 维乘积试验。

于是，T 维乘积试验给抽象的无穷维可测空间提供了一类现实的背景材料。

(2) 称 $\left(\prod_{t\in T}\Omega, \prod_{t\in T}\mathscr{F}\right)$ 为 $(\Omega,\mathscr{F})$ 的 T 重试验，简记为 $(\Omega^T,\mathscr{F}^T)$。现在，定理 2. 6. 3 可以推广到可列重试验。

例 7 T 维波莱尔试验 $(\mathbf{R}^T,\mathscr{B}^T)$ 是一维波莱尔试验 $(\mathbf{R},\mathscr{B})$ 的 T 重试验。

例 8 设 $(\Omega,\mathscr{F})$ 是贝努利试验(2. 4 节)。即

$$\Omega = \{A,\overline{A}\} \tag{2.6.40}$$

$$\mathscr{F} = \{\varnothing,\{A\},\{\overline{A}\},\Omega\} \tag{2.6.41}$$

那么可列重贝努利试验是 $(\Omega^\infty,\mathscr{F}^\infty)$，这里

$$\Omega^\infty = \{(\alpha_1,\alpha_2,\cdots,\alpha_i,\cdots)\mid \alpha_i = A \text{ 或 } \overline{A}, i = 1,2,3,\cdots\} \tag{2.6.42}$$

$$\mathscr{F}^\infty = \sigma(\mathscr{A}) \tag{2.6.43}$$

$$\mathscr{A} = \{\prod_{i=1}^{n} B_i \times \Omega \mid n \geqslant 1, B_i \in \mathscr{F}, i = 1,2,\cdots,n\} \tag{2.6.44}$$

2.6.7 注记

(1) 在联合试验的演算中，涉及样本空间的运算是集合论中

的运算，所以允许使用不可列并和不可列交。

(2) 我们用符号$\left(\underset{t\in T}{\circ}\Omega_t, \bigotimes_{t\in T}\mathscr{F}_t\right)$表示诸试验$(\Omega_t,\mathscr{F}_t)(t\in T)$的联合试验。新符号有简明的含义：$\underset{t\in T}{\circ}\Omega_t$是原因空间中的诸划分$\Omega_t(t\in T)$产生的交叉划分；$\bigotimes_{t\in T}\mathscr{F}_t$是包含诸$\mathscr{F}_t(t\in T)$的最小事件$\sigma$域。特别，当$T=\{1,2,\cdots,n\}$时它是$(\Omega_1\circ\Omega_2\circ\cdots\circ\Omega_n, \mathscr{F}_1\otimes\mathscr{F}_2\otimes\cdots\otimes\mathscr{F}_n)$；当$\underset{t\in T}{\circ}\Omega_t$是真交叉划分时它是乘积可测空间$\left(\prod_{t\in T}\Omega_t, \prod_{t\in T}\mathscr{F}_t\right)$。

(3) 把样本空间Ω_1加细成$\Omega_1\circ\Omega_2$时由(2.6.3)或(2.6.18)看出，Ω_1中的划分块$\omega_\alpha^{(1)}$被分裂成

$$\omega_\alpha^{(1)}=\bigcup_{\beta\in J_2}(\omega_\alpha^{(1)}\cap\omega_\beta^{(2)}) \tag{2.6.45}$$

即是说，Ω_1中的每个划分块皆用同样的方式分裂成$\Omega_1\circ\Omega_2$中的划分块。

一般地，我们可以把Ω_1中的一些划分块按一种方式分裂成更小的划分块；另一些划分块按另一种方式分裂成更小的划分块；…这样，可以得到比联合试验更复杂的试验。作者在构造全部右连续Q过程时曾使用过这种更复杂的产生随机试验的方法[20]。

2.7 第Ⅳ组公理：概率公理组

2.7.1 直观背景

毫无疑问，概率是概率论中最核心、最关键的一个概念。为了把它纳入自然公理系统，我们在1.2节中依据常识抽象出：

概率论原理Ⅰ：在合理划定(或给定)的条件$\mathscr{C}$下，涉及$\mathscr{C}$的事件A出现的可能性大小能够用[0,1]中惟一的实数——概率$P(A|\mathscr{C})$表达。

然后在1.3节中对概率进行了初步探讨。介绍了概率的4种

解释，以及多种解释并存的理论意义和实用价值。事实上，历史上的古典概率论和分析概率论就是建立在4种解释上的理论。

现在指出，4种解释隐含一个重大的缺点：它们仅仅把概率视为一个个的数值，忽略或没有强调它是在“划定条件”下的“数值族”。因而既不能把“数值族”作为整体进行逻辑推理和数学演算，又不能讨论“划定条件”和“数值族”的内在联系。

为了更清楚地表达这个缺点，不妨预先承认概率是一个特殊的函数(参看式(2.7.1))。那么4种解释的缺点是，它们仅仅给出确定某些函数值的方法(我们将把这些方法抽象为第Ⅵ组公理)，忽视了把这些函数值作为一个整体(即函数)的存在。因而很难把概率作为“函数”进行逻辑推理和数学演算，更难讨论“划定条件”与概率之间的内在联系。

克服这个缺点的方法是，完成1.2节交给自然公理系统的任务：把条件 $\mathscr{C}$，事件 A 和概率 $P(A|\mathscr{C})$ 抽象成能进行逻辑推理和数学演算的对象。

自然公理系统的前3组公理勾画出一幅情景：现实中的随机世界原来是一个因果空间。在这个空间中存在各种各样的“图形”——原因点，样本点，事件，事件域，事件 σ 域和随机试验。新奇的是，只有把“小图形”放在“大图形”中进行研究才能真正地理解这个“小图形”。例如，概率论原理Ⅰ现在被充实为：特定的事件存在于随机试验之中，只有认识该随机试验(认识的内容包括特定事件与试验中其他事件之间的联系)才能理解这个特定的事件。由此得出，“图形”的研究竟然是从随机试验开始，然后在这个试验中研究具体的事件和其他的“图形”。

概率论原理Ⅱ中划定的条件 $\mathscr{C}$ 在现实中被解释为某个实验 $\mathscr{E}$ 进行的条件。依据2.4节的讨论，实验 $\mathscr{E}$ 产生一个随机试验 $(\Omega,\mathscr{F})$。于是，原理Ⅱ蕴含 $\mathscr{F}$ 中的每个事件 A 出现的可能性大小皆是某实数值 $P(A|\mathscr{C})$。即是说，条件 $\mathscr{C}$ 产生一个定义在 $\mathscr{F}$

上的集函数

$$P(A|\mathscr{C}),A \in \mathscr{F} \qquad (2.7.1)$$

之所以称为集函数，是因为 $\mathscr{F}$ 中元素 A 是 Ω 的子集。

和"图形"的研究一样，概率的研究是从"集函数"，而不是从概率值开始。即是说，只有先认识了整体——"集函数"(即使无法确定它的数值，也应当知道它的定义域)，才能正确地理解概率值及其计算。论证这个观点的另一个初浅事实是，当某人作出判断"事件 A 出现的可能性是数值 p"时，不管他是否意识到，他同时也作出了判断"A 的对立事件出现的可能性是 $1-p$"，因而"判断"的内容实际上是(2.7.7)规定的集函数。

我们终于认识了概率的真面貌——它是随机试验$(\Omega,\mathscr{F})$上的集函数！既然 $\mathscr{F}$ 中的事件之间存在紧密的联系，反映这些事件出现可能性大小的数值族$\{P(A|\mathscr{C})|A\in\mathscr{F}\}$也必然相互依赖。这种依赖关系被 K 公理系统抽象成 3 条公理，即本节的概率公理组。

2.7.2 概率公理组

定义 2.7.1 设$(\Omega,\mathscr{F})$是随机试验，$P(A),A\in\mathscr{F}$ 是定义在 $\mathscr{F}$ 上的实值函数。称 $P(A),A\in\mathscr{F}$ 为$(\Omega,\mathscr{F})$上的概率测度，如果它满足如下 3 条公理：

公理IV_1(非负性公理) $P(A)\geqslant 0(A\in\mathscr{F})$；

公理IV_2(规范性公理) $P(\Omega)=1$；

公理IV_3(σ 可加性公理) 假定 $A_i(i\in I)$是 $\mathscr{F}$ 中有限或可列个互不相容的事件，则

$$P\Big(\bigcup_{i\in I} A_i\Big)=\sum_{i\in I} P(A_i) \qquad (2.7.2)$$

通常，概率测度被简称为概率，简记为 $P(\mathscr{F})$或 P[①]。

① 今后可以把任意的函数 $f(x),x\in\mathscr{D}$ 简记为 $f(\mathscr{D})$。

定义 2.7.2 称三元组$(\Omega,\mathscr{F},P)$为概率空间，如果$(\Omega,\mathscr{F})$是随机试验，P 是其上的概率测度。

定义 2.7.3 设$(\Omega_i,\mathscr{F}_i,P_i)(i=1,2)$是两个概率空间。称$(\Omega_2,\mathscr{F}_2,P_2)$是$(\Omega_1,\mathscr{F}_1,P_1)$的子概率空间，如果它们满足下列两个条件：

(1) $\mathscr{F}_2\subset\mathscr{F}_1$；

(2) 对任意的 $A\in\mathscr{F}_2$ 成立 $P_2(A)=P_1(A)$。

特别，$\Omega_1=\Omega_2$ 时称为同因子概率空间。

在正式展开对概率空间的研究之前，让我们给出定义 2.7.2 的直观意义。假定概率空间$(\Omega,\mathscr{F},P)$已经给定，那么我们有结论："在给定$(\Omega,\mathscr{F},P)$的条件下，$\mathscr{F}$ 中事件 A 出现的可能性大小是概率值 $P(A)$"。把它和原理 Ⅱ 对照后发现，概率空间$(\Omega,\mathscr{F},P)$全面地、准确地刻画出原理 Ⅱ 中的合理划定(或给定)的条件 $\mathscr{C}$。即

$$\text{划定的条件 } \mathscr{C}\Leftrightarrow(\Omega,\mathscr{F},P) \tag{2.7.3}$$

于是，条件 $\mathscr{C}$ 被抽象成$(\Omega,\mathscr{F},P)$，$\mathscr{C}$ 涉及的事件被抽象成 $\mathscr{F}$，出现可能性大小 $P(A|\mathscr{C})$被抽象成 $P(A)$，$A\in\mathscr{F}$，它们都是能够进行逻辑推理和数学演算的对象。因此，式(2.7.3)表明定义 2.7.2 严格地定义了条件 $\mathscr{C}$，概率记号(2.7.1)被简化为 $P(A)$，$A\in\mathscr{F}$。

在建模工作中，合理划定的条件 $\mathscr{C}$ 是实验 $\mathscr{E}$ 进行的条件。2.4节的讨论表明，实验 $\mathscr{E}$ 产生一个随机试验$(\Omega,\mathscr{F})$。现在，实验 $\mathscr{E}$ 还产生一个该试验上的概率测度 $P(A|\mathscr{C})$，$A\in\mathscr{F}$。即是说，实验 $\mathscr{E}$ 产生一个概率空间$(\Omega,\mathscr{F},P)$。把这个结论表示为

$$\text{实验 } \mathscr{E}(\text{在划定条件 } \mathscr{C} \text{ 下})\Rightarrow(\Omega,\mathscr{F},P) \tag{2.7.4}$$

显然，(2.7.4)构成现实随机世界中的一对因果关系，左方是起因，右方是后果。

今后，在演绎概率论理论体系时我们用"给定概率空间

$(\Omega,\mathscr{F},P)$”代替不精确的短语“在划定的条件 $\mathscr{C}$ 下进行实验 $\mathscr{E}$”(即使用式(2.7.3));在建模工作中,划定的条件 $\mathscr{C}$ 仍然是人人认可的、不言而喻的条件。人们将根据不言而喻的条件建立实验 $\mathscr{E}$ 产生的、合理的、和可接受的概率空间$(\Omega,\mathscr{F},P)$(即实现式(2.7.4),参看 2.4 节和 2.11 节)。

2.7.3 概率测度的存在性

第Ⅳ组公理首先面临的问题是:$P(\mathscr{F})$存在吗?惟一吗?柯尔莫哥洛夫回答:“我们的公理系统 Ⅰ～Ⅴ 是相容的(即无矛盾的)。这可以用下面的例子来验证:E 由一个元素 ξ 组成,$\mathscr{F}$ 由 E 与空集$\varnothing$组成,这里令 $P(E)=1,P(\varnothing)=0$①。”[1]接着,他列举出一个$P(\mathscr{F})$不惟一的例子,说明公理系统是不完备的。

菲纳特提出的第 7 个讨论题是:“上面提及的相容性证明令人满意吗?”[2]随后他表达了不满意的理由。我们把下面的引理作为这个讨论题的答案。

引理 2.7.1 对任意的随机试验$(\Omega,\mathscr{F})$,必定能够在其上赋予概率测度 $P(\mathscr{F})$。$P(\mathscr{F})$是惟一的当且仅当$(\Omega,\mathscr{F})$是退化试验。

证明 首先假定$(\Omega,\mathscr{F})$是退化试验。这时 $\mathscr{F}=\{\varnothing,\Omega\}$。令

$$P(A)=\begin{cases}0 & A=\varnothing\\ 1 & A=\Omega\end{cases} \tag{2.7.5}$$

显然 $P(\mathscr{F})$是概率测度,并且是惟一的。

然后假定$(\Omega,\mathscr{F})$是非退化的。这时存在 $B\in\mathscr{F}$ 使得 $B\neq\varnothing$ 且 $\overline{B}=\varnothing$。因此存在样本点 ω_1 和 ω_2 使得 $\omega_1\in B,\omega_2\in\overline{B}$。对任意的$A\in\mathscr{F}$,令

---

① 这个例子是(2.7.5)规定的退化的概率空间。

$$P(A)=\begin{cases}0 & \omega_1 \overline{\in} A \text{ 且 } \omega_2 \overline{\in} A \\ p & \omega_1 \in A \text{ 且 } \omega_2 \overline{\in} A \\ q & \omega_1 \overline{\in} A \text{ 且 } \omega_2 \in A \\ 1 & \omega_1 \in A \text{ 且 } \omega_2 \in A\end{cases} \tag{2.7.6}$$

其中 $p\geqslant 0, q\geqslant 0, p+q=1$。容易验证 $P(\mathscr{F})$是概率测度。$P(\mathscr{F})$的不惟一性由 p,q 的不惟一性推出。 **证毕**

例 1 设$(\Omega,\mathscr{F})$是(2.6.40)和(2.6.41)规定的伯努利试验。令

$$P(B)=\begin{cases}0 & B=\varnothing \\ p & B=A \\ q & B=\overline{A} \\ 1 & B=\Omega\end{cases} \tag{2.7.7}$$

其中 $p\geqslant 0, q\geqslant 0, p+q=1$。显然 $P(\mathscr{F})$是概率测度。

今后,称(2.7.5)规定的$(\Omega,\mathscr{F},P)$为退化概率空间;称例 1 中的$(\Omega,\mathscr{F},P)$为伯努利概率空间。

2.7.4 古典型概率空间

下面的引理表明概率的古典定义(1.3 节)满足第Ⅳ组公理。

引理 2.7.2 设$(\Omega,\mathscr{F})$是 n 型试验。对任意的 $A\in\mathscr{F}$,令

$$P(A)=\frac{A\text{ 含样本点的个数}}{n} \tag{2.7.8}$$

那么 $P(\mathscr{F})$是$(\Omega,\mathscr{F})$上的概率测度。

证明 $P(\mathscr{F})$显然满足公理Ⅳ_1 和Ⅳ_2。下证它满足公理Ⅳ_3。设 $A_1,A_2,\cdots,A_m$ 是互不相容的事件。由于诸 A_i 中没有公共的样本点,因此

$$P\left(\bigcup_{i=1}^{m} A_i\right)=\frac{\bigcup_{i=1}^{m} A_i\text{ 中含样本点的个数}}{n}$$

$$= \frac{A_1 \text{ 中含样本点的个数}}{n} + \cdots + \frac{A_m \text{ 中含样本点的个数}}{n},$$

$$= \sum_{i=1}^{m} P(A_i) \tag{2.7.9}$$

设 $A_i(i=1,2,3,\cdots)$ 是可列个互不相容的事件。由于 $\mathscr{F}$ 至多含 2^n 个元素，因此诸 A_i 中非空事件为有限个，不妨设它们是 $A_1, A_2, \cdots, A_m$。注意到 $A_i = \varnothing(i \geqslant m+1)$，(2.7.8)含有 $P(\varnothing)=0$，故

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = P\Big(\bigcup_{i=1}^{m} A_i\Big) = \sum_{i=1}^{m} P(A_i) = \sum_{i=1}^{\infty} P(A_i)$$

得证 $P(\mathscr{F})$ 满足 σ 可加性公理。 **证毕**

推论 设$(\Omega, \mathscr{F})$是 n 型标准试验，$P(\mathscr{F})$是概率测度。那么(2.7.8)成立当且仅当对任意的 $\omega \in \Omega$ 有

$$P(\{\omega\}) = \frac{1}{n} \tag{2.7.10}$$

其中$\{\omega\}$是由单个样本点 ω 组成的集合。

定义 2.7.4 称引理 2.7.2 中的$(\Omega, \mathscr{F}, P)$为古典型概率空间，且称 $P(\mathscr{F})$为古典型概率测度。

2.7.5 几何型概率空间

下面的引理表明几何概率(1.3 节)满足第Ⅳ组公理。用 $\mu(\mathscr{B}^n)$ 表示 n 维波莱尔试验$(\mathbf{R}^n, \mathscr{B}^n)$上的勒贝格(H. L. Lebesgue)测度。众所周知，$\mu(\mathscr{B})$是长度；$\mu(\mathscr{B}^2)$是面积；$\mu(\mathscr{B}^3)$是体积。

引理 2.7.3 设$(D, \mathscr{B}^n(D))$是 D 上的波莱尔试验(定义 2.5.7)，$0 < \mu(D) < +\infty$。对任意的 $A \in \mathscr{B}^n(D)$，令

$$P(A) = \frac{\mu(A)}{\mu(D)} \tag{2.7.11}$$

那么 P 是$(D, \mathscr{B}^n(D))$上的概率测度。

证明 P 显然满足公理 IV_1 和 IV_2。应用勒贝格测度的 σ 可加性立即推出 P 满足公理 IV_3。 **证毕**

定义 2.7.5 称引理 2.7.3 中的$(D,\mathscr{B}^n(D),P)$为几何型概率空间，且称 P 为几何概率测度。

古典型和几何型概率空间在概率论的早期发展中起着基本的作用。这里不准备给出具体的例子，因为在初等概率论的书籍中能找到大量的例子。至于这些例子在建模时为什么产生古典型和几何型概率空间，我们将在第 Ⅵ 组公理(2.11 节)中讨论。

2.7.6 概率的频率解释和主观解释是否满足第Ⅳ组公理？

为了叙述流畅，我们首先引进两条公理。设$(\Omega,\mathscr{F})$是随机试验，$P(A)$，$A\in\mathscr{F}$ 是实值函数。

公理 IV_4(有限可加性公理) 设 A 和 B 是 $\mathscr{F}$ 中不相容的事件，则

$$P(A\cup B)=P(A)+P(B) \qquad (2.7.12)$$

公理 IV_5(连续性公理) 设 $A_i\in\mathscr{F}(i\in I)$，并且 $A_1\supset A_2\supset\cdots\supset A_i\supset\cdots$，$\bigcap_{i=1}^{\infty}A_i=\varnothing$。则

$$\lim_{i\to\infty}P(A_i)=0 \qquad (2.7.13)$$

然后介绍本节即将证明的两个结论：

结论 1(定理 2.7.1(6)) 设 $A_i(i=1,2,\cdots,n)$是 $\mathscr{F}$ 中互不相容的事件，那么(2.7.12)等价于

$$P\Big(\bigcup_{i=1}^{n}A_i\Big)=\sum_{i=1}^{n}P(A_i) \qquad (2.7.14)$$

结论 2(定理 2.7.3) 设$(\Omega,\mathscr{F})$是随机试验。又设 $P(A)$，$A\in\mathscr{F}$ 是实值集函数，并且满足公理 IV_1，IV_2 和 IV_4。那么 $P(\mathscr{F})$满足公理 IV_3 的必要和充分条件是它满足公理 IV_5。

由(1.3.14)和(1.3.15)看出概率的频率解释满足公理Ⅳ$_1$，Ⅳ$_2$和Ⅳ$_4$(从而也满足结论1)，且不排斥公理Ⅳ$_3$(或公理Ⅳ$_5$)。

同样，概率的主观解释显然遵守公理Ⅳ$_1$和Ⅳ$_2$。主观解释的一个信条是：要求判断前后一致，即相容性或相互无矛盾性。相容性要求导致遵守公理Ⅳ$_4$(从而遵守结论1)，也不排斥公理Ⅳ$_3$(或公理Ⅳ$_5$)[2]。

自然公理系统认为，有限和可列处于同样的地位。既然频率解释和主观解释都支持公理Ⅳ$_4$，因此也认为这两种解释满足公理Ⅳ$_3$(或公理Ⅳ$_5$)。

目前的实际情况是：菲纳特等一批学者不支持公理Ⅳ$_3$，认为应当代之为公理Ⅳ$_4$[2,8]。为了反映他们的观点，我们给出

定义 2.7.6 设$(\Omega,\mathscr{F})$是随机试验，$P(A)$，$A\in\mathscr{F}$是定义在$\mathscr{F}$上的实值函数。称$P(A)$，$A\in\mathscr{F}$是有限可加概率，如果它满足公理Ⅳ$_1$，Ⅳ$_2$和Ⅳ$_4$。

与此相区别，不妨称定义2.7.1中的P为σ可加概率。显然，σ可加概率是有限可加概率，反之则未必成立。

2.7.7 概率测度的基本性质

定理 2.7.1 设$(\Omega,\mathscr{F})$是随机试验，$P(\mathscr{F})$是有限可加概率。那么$P(\mathscr{F})$具有下列性质：

(1) $P(\varnothing)=0$；

(2) $0\leqslant P(A)\leqslant 1(A\in\mathscr{F})$；

(3) 互补性：$P(A)+P(\overline{A})=1(A\in\mathscr{F})$；

(4) 单调性：当$A\subset B$时有$P(A)\leqslant P(B)$；

(5) 包含可减性：当$A\subset B$时有$P(B\backslash A)=P(B)-P(A)$；

(6) 有限可加性：设$A_1,A_2,\cdots,A_n(n\geqslant 2)$是互不相容的事件，则成立(2.7.14)；

(7) 半可加性：设$A_1,A_2,\cdots,A_n(n\geqslant 2)$是$\mathscr{F}$中任意的事件，

则成立

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} P(A_i) \tag{2.7.15}$$

(8) 一般加法公式：设 $A_1, A_2, \cdots, A_n (n \geqslant 2)$ 是 $\mathscr{F}$ 中任意的事件，则

$$P\left(\bigcup_{i=1}^{n} A_i\right) = s_1 - s_2 + s_3 - \cdots + (-1)^{n-1} s_n \tag{2.7.16}$$

其中 $s_1 = \sum_{i=1}^{n} P(A_i)$，$s_2 = \sum_{i<j\leqslant n} P(A_i A_j)$，$s_3 = \sum_{i<j<k\leqslant n} P(A_i A_j A_k)$，$\cdots$，$s_n = P(A_1 A_2 \cdots A_n)$。

证明 (1) 由于 $\varnothing = \varnothing \cup \varnothing$，由公理 IV_4 得出 $P(\varnothing) = P(\varnothing) + P(\varnothing)$。得证 $P(\varnothing) = 0$。

(3) 由于 $A \cup \overline{A} = \Omega$，应用公理 IV_4 和 IV_2 即得结论。

(2) 应用(3)和公理 IV_1 即得结论。

(5) 设 $A \subset B$。容易验证

$$A \cup (B \backslash A) = B, B \cap (B \backslash A) = \varnothing \tag{2.7.17}$$

由公理 IV_4 得出 $P(A) + P(B \backslash A) = P(B)$。得证包含可减性结论。

(4) 应用(5)和公理 IV_1 即得结论。

(6) 注意到 A_n 和 $\bigcup_{i=1}^{n-1} A_i (n \geqslant 2)$ 是不相容的事件，应用公理 IV_4 得

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n)$$

对右方第一项继续进行这样的操作，共进行 $n-1$ 次操作后即得(2.7.14)。

(7) 先证 $n=2$ 的情形。显然有

$$A_1 \cup A_2 = A_1 \cup (A_2 \backslash A_1 A_2),\ A_1 \cap (A_2 \backslash A_1 A_2) = \varnothing \tag{2.7.18}$$

应用公理IV_4和概率的单调性得

$$P(A_1 \cup A_2) = P(A_1) + P(A_2 \backslash A_1 A_2) \leqslant P(A_1) + P(A_2) \tag{2.7.19}$$

现在假定(2.7.15)在$n \leqslant r$时成立，我们有

$$P\Big(\bigcup_{i=1}^{r+1} A_i\Big) \leqslant P\Big(\bigcup_{i=1}^{r} A_i\Big) + P(A_{r+1}) \leqslant \sum_{i=1}^{r+1} P(A_i)$$

由归纳法推出(2.7.15)对任意的正整数n成立。

(8) 对(2.7.19)中的等式部分应用概率的包含可减性得

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 A_2) \tag{2.7.20}$$

得证(2.7.16)在$n=2$时成立。假定(2.7.16)在$n \leqslant r$时成立，因此

$$P\Big(\bigcup_{i=1}^{r} A_i\Big) = \sum_{i=1}^{r} P(A_i) - \sum_{i<j\leqslant r} P(A_i A_j) + \cdots + (-1)^r P(A_1 A_2 \cdots A_r)$$

$$P\Big[\bigcup_{i=1}^{r} (A_i A_{r+1})\Big] = \sum_{i=1}^{r} P(A_i A_{r+1}) - \sum_{i<j\leqslant r} P(A_i A_j A_{r+1}) + \cdots + (-1)^r P(A_1 A_2 \cdots A_r A_{r+1})$$

另一方面，由(2.7.20)推出

$$P\Big(\bigcup_{i=1}^{r+1} A_i\Big) = P\Big(\bigcup_{i=1}^{r} A_i\Big) + P(A_{r+1}) - P\Big[\bigcup_{i=1}^{r} (A_i A_{r+1})\Big]$$

把前面的两式代入其中，经过简单的整理即得$n=r+1$时的(2.7.16)。 **证毕**

定理 2.7.2(连续性定理) 设$(\Omega, \mathscr{F})$是随机试验，$P(\mathscr{F})$是σ可加概率测度。那么$P(\mathscr{F})$具有下列两个性质：

(1) 下连续性：设$A_i \in \mathscr{F}(i \in I)$，$A_1 \subset A_2 \subset \cdots \subset A_i \subset \cdots$。令$A = \bigcup_{i=1}^{\infty} A_i$，则

$$P(A) = \lim_{i\to\infty} P(A_i) \tag{2.7.21}$$

(2) 上连续性：设$A_i \in \mathscr{F}(i \in I)$，$A_1 \supset A_2 \supset \cdots \supset A_i \supset \cdots$。

令$A=\bigcap\limits_{i=1}^{\infty}A_i$，则

$$P(A)=\lim_{i\to\infty}P(A_i) \tag{2.7.22}$$

证明 (1) 令 $A_0=\varnothing$，$B_i=A_i\setminus A_{i-1}(i=1,2,3,\cdots)$。容易验证诸$B_i(i\in I)$互不相容，并且

$$A=\bigcup_{i=1}^{\infty}A_i=\bigcup_{i=1}^{\infty}B_i \tag{2.7.23}$$

应用公理IV_3和概率的包含可减性得

$$\begin{aligned}P(A)&=\sum_{i=1}^{\infty}P(B_i)=\sum_{i=1}^{\infty}[P(A_i)-P(A_{i-1})]\\&=\lim_{i\to\infty}P(A_i)\end{aligned}$$

(2) 令 $B_i=\overline{A}_i(i=1,2,3,\cdots)$。由引理 2.2.3 得出 B_1,B_2,B_3，$\cdots$是满足(1)中条件的序列，并且 $\overline{A}=\bigcup\limits_{i=1}^{\infty}B_i$(德·摩根律)。应用(2.7.21)得

$$P(\overline{A})=\lim_{i\to\infty}P(\overline{A}_i) \tag{2.7.24}$$

注意到 $P(\overline{A})=1-P(A)$、$P(\overline{A}_i)=1-P(A_i)$，代入其中即得(2.7.22)。 **证毕**

定理 2.7.3 设$(\Omega,\mathscr{F})$是随机试验。那么$P(\mathscr{F})$是σ可加概率测度的充分必要条件是$P(\mathscr{F})$是有限可加概率，且满足公理IV_5。

证明 必要性由σ可加概率的上连续性推出。下证充分性。设$A_i(i\in I)$是公理IV_3中的事件，令

$$A=\bigcup_{i=1}^{\infty}A_i,\quad B_n=\bigcup_{i=1}^{n}A_i,\quad C_n=\bigcup_{i=n+1}^{\infty}A_i \tag{2.7.25}$$

显然B_n和C_n互不相容，且$A=B_n\cup C_n$。由于$P(\mathscr{F})$是有限可加概率，故可应用公理IV_4和(2.7.14)得到

$$P(A)=P(B_n)+P(C_n)$$

$$= \sum_{i=1}^{n} P(A_i) + P(C_n) \tag{2.7.26}$$

令 $n \to \infty$ 得 $P(A) = \sum_{i=1}^{\infty} P(A_i) + \lim_{n\to\infty} P(C_n)$。为完成证明，只须证 $\lim_{n\to\infty} P(C_n) = 0$。

应用 $A_i (i \in I)$ 的不相容性得出 $C_1 \supset C_2 \supset \cdots \supset C_n \supset \cdots$。由 $C_n = A \setminus B_n = A\overline{B}_n$ 推出 $\overline{C}_n = \overline{A} \cup B_n$，故

$$\overline{\bigcap_{n=1}^{\infty} C_n} = \bigcup_{n=1}^{\infty} \overline{C}_n = \overline{A} \cup \left(\bigcup_{n=1}^{\infty} B_n \right) = \overline{A} \cup A = \Omega$$

即 $\bigcap_{n=1}^{\infty} C_n = \varnothing$。因此 $C_1, C_2, \cdots, C_n, \cdots$ 是满足公理 IV_5 中条件的事件序列。应用这个公理得证 $\lim_{n\to\infty} P(C_n) = 0$。 **证毕**

2.7.8 概率测度的扩张

能够像(2.7.8)和(2.7.11)那样"整体"定义的概率测度是很少的。通常的情形是：设 $\mathscr{F} = \sigma(\mathscr{A})$，我们能够在芽事件集 $\mathscr{A}$ 上赋予概率值

$$P(A), A \in \mathscr{A} \tag{2.7.27}$$

试问 $\mathscr{A}$ 和 $P(\mathscr{A})$ 应当满足什么样的条件才能在 $\mathscr{F}$ 上惟一地确定一个概率测度 $P(\mathscr{F})$，使得 $P(\mathscr{F})$ 在 $\mathscr{A}$ 上的值是(2.7.27)？简言之，在什么样的条件下能够把 $P(\mathscr{A})$ 扩张成 $P(\mathscr{F})$？

实际上，扩张问题是测度论的基本问题之一。为了叙述这个问题，需要引进

定义 2.7.7 设 $\mathscr{R}$ 是因果空间 $(X, \mathscr{U})$ 中的事件域，$P(A)$，$A \in \mathscr{R}$ 是定义在 $\mathscr{R}$ 上的实值函数。称 $P(A), A \in \mathscr{R}$ 为 $\mathscr{R}$ 上的概率测度，如果它满足公理 IV_1，公理 IV_2 和条件：假定 $A_i (i \in I)$ 是 $\mathscr{R}$ 中有限或可列个互不相容的事件，且 $\bigcup_{i \in I} A_i \in \mathscr{R}$，则

$$P\left(\bigcup_{i\in I}A_i\right)=\sum_{i\in I}P(A_i) \tag{2.7.28}$$

定理 2.7.4(概率测度扩张定理) 设$(\Omega,\mathscr{F})$是随机试验,$\mathscr{R}$是Ω上的事件域,且$\mathscr{F}=\sigma(\mathscr{R})$。如果$P(\mathscr{R})$是$\mathscr{R}$上的概率测度,那么存在$\mathscr{F}$上的概率测度$\overline{P}(\mathscr{F})$,满足

$$\overline{P}(A)=P(A),\text{当 } A\in\mathscr{R} \tag{2.7.29}$$

并且$\overline{P}(\mathscr{F})$是惟一的。

这个定理是测度论中测度扩张定理的特殊情形,证明见文献[16]。通常,把$\overline{P}(\mathscr{F})$仍记为$P(\mathscr{F})$,并称$P(\mathscr{F})$是$P(\mathscr{R})$的扩张。

定理 2.7.5 设$(\underset{t\in T}{\circ}\Omega_t,\bigotimes_{t\in T}\mathscr{F}_t,P)$是概率空间,其中$(\underset{t\in T}{\circ}\Omega_t,\bigotimes_{t\in T}\mathscr{F}_t)$是$(\Omega_t,\mathscr{F}_t)(t\in T)$的联合随机试验。那么概率测度$P(\bigotimes_{t\in T}\mathscr{F}_t)$由它的部分数值

$$P(A),A\in\underset{t\in T}{*}\mathscr{F}_t \tag{2.7.30}$$

惟一地确定。这里$\underset{t\in T}{*}\mathscr{F}_t$是(2.6.38)规定的柱事件集。

证明 令

$$\mathscr{R}=\left\{\bigcup_{i=1}^{m}B_i\,|\,m\geqslant 1,B_1,B_2,\cdots,B_m\text{ 是互不相容的柱事件}\right\} \tag{2.7.31}$$

先证$\mathscr{R}$关于补封闭。设$B=A_{t_1}\cap A_{t_2}\cap\Omega_t$是一个二维柱事件,我们有

$$\overline{B}=(A_{t_1}\cap\overline{A}_{t_2}\cap\Omega_t)\cup(\overline{A}_{t_1}\cap A_{t_2}\cap\Omega_t)\cup(\overline{A}_{t_1}\cap\overline{A}_{t_2}\cap\Omega_t) \tag{2.7.32}$$

显然右方是三个不相容的柱事件,故$\overline{B}\in\mathscr{R}$。类似地可证对任意的$n$维柱事件$B$成立$\overline{B}\in\mathscr{R}$。

又设$B=\bigcup_{i=1}^{m}B_i$是$\mathscr{R}$中任意元素,即$B_1,B_2,\cdots,B_m$是互不相容的柱事件。由德·摩根律得

$$\overline{B}=\bigcap_{i=1}^{m}\overline{B}_i \tag{2.7.33}$$

由已证的结论得出 $\overline{B}_i$ 是不相容事件的有限并。对上式右方应用并和交的分配律(定理 2.2.3)得出,B 是一些不相交事件的有限并,其中的每一个加项皆是有限个柱事件的交,从而它是柱事件(引理 2.6.1(2)),故 $\overline{B}\in\mathscr{R}$。得证 $\mathscr{R}$ 关于补封闭。类似地可证 $\mathscr{R}$ 关于差封闭。

显然,$\mathscr{R}$ 关于不相容事件的有限并封闭。应用 2.2 节末注记(2)即得 $\mathscr{R}$ 关于有限并封闭。得证 $\mathscr{R}$ 是事件域。显然,$\sigma(\mathscr{R})=\bigotimes\limits_{t\in T}\mathscr{F}_t$。

最后,由(2.7.14)得出数值族(2.7.30)惟一地确定$P(\mathscr{R})$。再应用定理 2.7.4 得出它惟一地确定概率测度 $P(\bigotimes\limits_{t\in T}\mathscr{F}_t)$。 **证毕**

注意,$\mathop{*}\limits_{t\in T}\mathscr{F}_t$ 中不同形式的表达式可以代表同一个事件。因此,如果(2.7.30)的数值不是预先给定,而是进行赋值,那么为了保证 $P(\mathop{*}\limits_{t\in T}\mathscr{F}_t)$ 的单值性必须掌握 $\mathop{*}\limits_{t\in T}\mathscr{F}_t$ 的结构。引理2.6.2的重要性在于显示乘积随机试验的 $\mathop{*}\limits_{t\in T}\mathscr{F}_t$ 有简单的、清楚的结构。这时只有不可能事件才存在不同形式的表达式,因而下面的(2.7.34)规定的 $P(\mathop{*}\limits_{t\in T}\mathscr{F}_t)$ 显然是单值函数。

2.7.9 独立乘积概率空间

定理 2.7.6 设$(\Omega_t,\mathscr{F}_t,P_t)(t\in T)$ 是一族概率空间。又设$(\Omega_t,\mathscr{F}_t)(t\in T)$ 的联合试验是乘积试验$\left(\prod\limits_{t\in T}\Omega_t,\prod\limits_{t\in T}\mathscr{F}_t\right)$。对形如(2.6.37)的柱事件赋予概率值

$$P\left(\prod_{i=1}^{n}A_{t_i}\times\prod\Omega_t\right)=\prod_{i=1}^{n}P_{t_i}(A_{t_i}) \tag{2.7.34}$$

那么 $P(\mathop{*}\limits_{t\in T}\mathscr{F}_t)$ 可以惟一地扩张成概率测度 $P\left(\prod\limits_{t\in T}\mathscr{F}_t\right)$。

证明见文献[16]。

定义 2.7.8 称上述定理中的 $P\left(\prod_{t\in T}\mathscr{F}_t\right)$ 为概率族 $P_t(\mathscr{F}_t)(t\in T)$ 的独立乘积概率测度，简记为 $\prod_{t\in T}P_t$，且称 $\left(\prod_{t\in T}\Omega_t,\prod_{t\in T}\mathscr{F}_t,\prod_{t\in T}\mathscr{F}_t\right)$ 为 $(\Omega_t,\mathscr{F}_t,P_t)(t\in T)$ 的独立乘积概率空间。

特别，当 $(\Omega_t,\mathscr{F}_t,P_t)$ 不依赖于 t 时称它为 $(\Omega,\mathscr{F},P)$ 的独立重复概率空间，记为 $(\Omega^T,\mathscr{F}^T,P^T)$。

容易看出，在乘积概率空间 $\left(\prod_{t\in T}\Omega_t,\prod_{t\in T}\mathscr{F}_t,\prod_{t\in T}P_t\right)$ 中 $(\Omega_t,\mathscr{F}_t,P_t)$ 是它的子概率空间，$\left(\prod_{t\in T}\Omega_t,\mathscr{F}_t,P_t\right)$ 是它的同因子概率空间。

为了和 2.7.1 段呼应，让我们指出此处概率的演算和 2.7.7 段中概率的演算有本质的差别。后者是在固定一个概率测度的前提下讨论概率值之间的计算，而定理 2.7.6 是一族概率测度产生出新概率测度的演算。

例 2 设 $(\Omega,\mathscr{F},P)$ 是例 1 中的伯努利概率空间。称 $(\Omega^\infty,\mathscr{F}^\infty,P^\infty)$ 为可列重伯努利概率空间，其中 $(\Omega^\infty,\mathscr{F}^\infty)$ 由(2.6.42)～(2.6.44)规定，P^∞ 是独立乘积概率测度。对(2.6.42)中的 $\omega\overset{\text{def}}{=}(\alpha_1,\alpha_2,\cdots,\alpha_i,\cdots)$，定义

$$\nu_{nA}(\omega)=\alpha_1,\alpha_2,\cdots,\alpha_n\text{ 中包含事件 }A\text{ 的个数} \tag{2.7.35}$$

显然，对固定的 n，集合

$$\left\{\left|\frac{\nu_{nA}(\omega)}{n}-p\right|<\varepsilon\right\}\overset{\text{def}}{=}\left\{\omega\left|\left|\frac{\nu_{nA}(\omega)}{n}-p\right|<\varepsilon\right.\right\} \tag{2.7.36}$$

是(2.6.44)中柱事件的有限并。由此推出

$$\left\{\lim_{n\to\infty}\frac{\nu_{nA}(\omega)}{n}=p\right\}\overset{\text{def}}{=}\left\{\omega\left|\lim_{n\to\infty}\frac{\nu_{nA}(\omega)}{n}=p\right.\right\} \tag{2.7.37}$$

是$\mathscr{F}^{\infty}$中的事件。我们有

定理 2.7.7(伯努利大数定理)　在可列重伯努利概率空间$(\Omega^{\infty},\mathscr{F}^{\infty},P^{\infty})$中，对任意的$\varepsilon(>0)$成立

$$\lim_{n\to\infty}P^{\infty}\left(\left|\frac{\nu_{nA}(\omega)}{n}-p\right|<\varepsilon\right)=1 \tag{2.7.38}$$

定理 2.7.8(波莱尔大数定理)　在可列重伯努利概率空间$(\Omega^{\infty},\mathscr{F}^{\infty},P^{\infty})$中成立

$$P^{\infty}\left(\lim_{n\to\infty}\frac{\nu_{nA}(\omega)}{n}=p\right)=1 \tag{2.7.39}$$

为了节省篇幅，略去证明。适合此处需要的初等证明见文献[21]。这两个定理为概率的频率解释(1.3 节)提供科学的依据。

2.7.10 注记

(1) 设$(\Omega,\mathscr{F},P)$是概率空间。由于零概率集的子集不一定是事件；即使是事件，也不一定属于$\mathscr{F}$。这个事实给研究工作带来不必要的麻烦。消除这个麻烦的方法是，认为零概率集的任何子集皆是事件，且属于$\mathscr{F}$，其概率为零。表达方式为

$$\overline{\mathscr{F}}=\{A\Delta N\mid A\in\mathscr{F},N\text{ 为零概率集的子集}\} \tag{2.7.40}$$

$$\overline{P}(A\Delta N)=P(A) \tag{2.7.41}$$

其中Δ是 2.2 节末注记 3)引进的对称差。习惯上，把$\overline{P}(\overline{\mathscr{F}})$仍记为$P(\overline{\mathscr{F}})$。容易验证$(\Omega,\overline{\mathscr{F}},P)$是概率空间，且$(\Omega,\mathscr{F},P)$是它的同因子概率空间。

称$(\Omega,\overline{\mathscr{F}},P)$为$(\Omega,\mathscr{F},P)$的增补概率空间。称$P(\overline{\mathscr{F}})$为$P(\mathscr{F})$的增补，或增补概率测度。

(2) 设$(\Omega,\mathscr{F},P)$是概率空间，那么$\mathscr{F}$中至多存在可列个不相容的原子事件$B_i(i\in I)$使得$P(B_i)>0$。事实上，概率大于$\frac{1}{n}$的原子事件至多$n-1$个，故概率大于零的原子事件至多可列个。

用 $B_i(i\in I)$表示概率大于零的全体原子事件，令

$$\Omega' = \left\{B_i,\omega \mid i\in I,\omega\in \Omega\setminus \bigcup_{i\in I} B_i\right\} \qquad (2.7.42)$$

容易证明 Ω'是 Ω 的加粗，$(\Omega',\mathscr{F},P)$是概率空间。显然增补概率空间$(\Omega',\overline{\mathscr{F}},P)$中的$(\Omega',\overline{\mathscr{F}})$是标准随机试验。于是，增补方法完成了 2.5 节末注记中提及的标准化任务，且成为流行的方法。

(3) 在采用有限可加概率和 σ 可加概率的争论中，作者赞成菲什伯恩(P. C. Fishburn)的观点："只用有限可加概率就能得到结论是令人向往的。但是，由于拒绝公理 N_3 或 N_5(即本节的公理Ⅳ$_3$ 或Ⅳ$_5$)得不到结论，或者引起不必要的复杂性，那么我们就毫不犹豫地引用公理 N_3 或 N_5。"[8]

(4) 自然公理系统对上面的争论采取兼收并蓄的态度。认为有限可加和 σ 可加概率空间都是因果空间中的"图形"，且鼓励引进更多的"图形"。重要的是发现"图形"中蕴含的统计思想和统计规律，以及"图形"之间的联系。

2.8 随机试验上的点函数

在概率空间$(\Omega,\mathscr{F},P)$中 $\mathscr{F}$ 是我们关心的全部事件组成的集合，$P(\mathscr{F})$是这些事件出现可能性的度量。因此$(\mathscr{F},P)$是概率空间的主体。第Ⅰ组公理说，Ω 是认识 $\mathscr{F}$ 的重要工具。本节论证 Ω 也是认识 $P(\mathscr{F})$的重要工具。

随机试验$(\Omega,\mathscr{F})$上存在两类函数。一类是定义在 $\mathscr{F}$ 上的函数，称为集函数。概率测度是一类特殊的集函数。另一类是定义在 Ω 上的函数，称为点函数。一般说来，点函数比集函数简单。

本节的任务是介绍几类众所周知的点函数。应用它们能够确定典型随机试验上的所有概率测度，从而得到几类典型的概率空间。它们在概率论的理论体系中起着核心作用。

2.8.1 离散型概率空间

定义 2.8.1 称概率空间$(\Omega,\mathscr{F},P)$是离散型的(或 n 型的,或可列型的),如果随机试验$(\Omega,\mathscr{F})$是离散型的(或 n 型的,或可列型的)。

定义 2.8.2 设$(\Omega,\mathscr{F})$是离散型试验,$f(\omega),\omega\in\Omega$ 是实值函数。称 $f(\omega),\omega\in\Omega$ 为$(\Omega,\mathscr{F})$上的分布律,或概率分布律,如果它是非负的,且 $\sum\limits_{\omega\in\Omega}f(\omega)=1$。

离散型试验的样本点可以一一列举。不失一般性,不妨设

$$\Omega=\{\omega_1,\omega_2,\cdots,\omega_i,\cdots\}=\{\omega_i|i\in I\} \tag{2.8.1}$$

它可以是 n 元集,也可以是可列集。这时分布律 $f(\omega),\omega\in\Omega$ 也可采用函数的表格法表示为

$$f:\begin{bmatrix}\omega_1 & \omega_2 & \cdots & \omega_i & \cdots \\ f(\omega_1) & f(\omega_2) & \cdots & f(\omega_i) & \cdots\end{bmatrix}\ \left(\sum_{i\in I}f(\omega_i)=1\right) \tag{2.8.2}$$

因此,又称 $f(\omega)$为分布列,或概率分布列。

定理 2.8.1 设$(\Omega,\mathscr{F})$是离散型试验。那么分布律 $f(\omega)$用公式

$$P(A)=\sum_{\omega\in A}f(\omega)\quad (A\in\mathscr{F}) \tag{2.8.3}$$

确定一个概率测度 $P(\mathscr{F})$。反之,设 $P(\mathscr{F})$是概率测度,那么存在分布律 $f(\omega)$使得(2.8.3)成立,这个 $f(\omega)$是惟一的当且仅当$(\Omega,\mathscr{F})$中不存在原子事件 B 使得 $P(B)>0$,且 B 至少包含两个样本点。

证明 设 $f(\omega)$是分布律,$P(\mathscr{F})$用(2.8.3)规定。显然 $P(\mathscr{F})$满足公理Ⅳ$_1$和Ⅳ$_2$。假定 $A_1,A_2,\cdots,A_k,\cdots$是 $\mathscr{F}$ 中有限或可列个互不相容的事件,不妨设

$$A_k=\{\omega_i|i\in I_k\}\quad (k=1,2,3,\cdots)$$

其中 I_k 是 I 的子集(见(2.8.1))。显然诸 $I_k(k\geqslant 1)$ 互不相交。令 $I_0=\bigcup\limits_{k\geqslant 1} I_k$，应用收敛正项级数$\left(因为\sum\limits_{i\in I_0} f(\omega_i)\leqslant 1\right)$的交换律和结合律，由(2.8.3)得出

$$P\left(\bigcup_{k\geqslant 1} A_k\right)=\sum_{i\in I_0} f(\omega_i)=\sum_{k\geqslant 1}\left[\sum_{i\in I_k} f(\omega_i)\right]=\sum_{k\geqslant 1} P(A_k)$$

故 $P(\mathscr{F})$满足公理Ⅳ$_3$，得证定理的前半部分。

下证“反之”部分。由定理 2.5.3 和定理 2.5.6 知，存在有限或可列个互不相容的原子事件 $B_k(k=1,2,3,\cdots)$使得

$$\Omega=\bigcup_{k\geqslant 1} B_k \tag{2.8.4}$$

显然存在非负的实值函数 $f(\omega)$满足

$$P(B_k)=\sum_{\omega\in B_k} f(\omega) \qquad (k=1,2,3,\cdots) \tag{2.8.5}$$

由于

$$\sum_{\omega\in\Omega} f(\omega)=\sum_{k\geqslant 1}\left[\sum_{\omega\in B_k} f(\omega)\right]=\sum_{k\geqslant 1} P(B_k)=1$$

故 $f(\omega)$是分布律。对任意的 $A\in\mathscr{F}$，由(2.8.4)得

$$A=\bigcup_{k\geqslant 1}(AB_k) \tag{2.8.6}$$

由于右方诸事件互不相容，由 $P(\mathscr{F})$的 σ 可加性得

$$P(A)=\sum_{k\geqslant 1} P(AB_k)$$

注意到 AB_k 或为空集，或为 B_k，应用(2.8.4)得

$$P(A)=\sum_{k\geqslant 1}\left[\sum_{\omega\in AB_k} f(\omega)\right]=\sum_{\omega\in A} f(\omega)$$

得证(2.8.3)成立。剩下证惟一性的结论。

如果存在原子事件 B 使得 $P(B)>0$，且 B 至少包含两个样本点，显然满足(2.8.5)的 $f(\omega)$有无穷多个，得证 $f(\omega)$的不惟一性。如果这样的 B 不存在，满足(2.8.5)的 $f(\omega)$存在且惟一，它是

$$f:\begin{pmatrix} \omega_1 & \omega_2 & \cdots & \omega_i & \cdots \\ P(\{\omega_1\}) & P(\{\omega_2\}) & \cdots & P(\{\omega_i\}) & \cdots \end{pmatrix} \tag{2.8.7}$$

其中约定，当$\langle\omega_i\rangle\overline{\in}\mathscr{F}$时认为$P(\langle\omega_i\rangle)=0$。 **证毕**

推论 当$(\Omega,\mathscr{F})$是标准试验时，$P(\mathscr{F})$和分布律(2.8.7)相互惟一确定。

定义 2.8.3 称(2.8.5)规定的点函数$f(\omega)$为概率测度$P(\mathscr{F})$的分布律。

2.8.2 柯尔莫哥洛夫概率空间[①]

定义 2.8.4 称概率空间$(\mathbf{R},\mathscr{B},P)$为柯尔莫哥洛夫的，如果$(\mathbf{R},\mathscr{B})$是一维波莱尔试验。

定义 2.8.5 称定义在实数集$\mathbf{R}$上的非负实值函数$F(x)$，$x\in\mathbf{R}$为分布函数，如果$F(x)$具有下列3个性质：

(1) 单调不减性：$F(x_1)\leqslant F(x_2)$，当$x_1<x_2$；

(2) 右连续性：$\lim\limits_{x\to a^+}F(x)=F(a)$（对任意的$a\in\mathbf{R}$）；

(3) 规范性：

$$\left.\begin{aligned}F(-\infty)&\overset{\text{def}}{=\!=}\lim_{x\to-\infty}F(x)=0\\F(+\infty)&\overset{\text{def}}{=\!=}\lim_{x\to+\infty}F(x)=1\end{aligned}\right\}\tag{2.8.8}$$

定理 2.8.2 设$(\mathbf{R},\mathscr{B})$是一维波莱尔试验。那么分布函数$F(x)$用公式

$$P((a,b])=F(b)-F(a)\qquad(-\infty\leqslant a<b\leqslant+\infty)\tag{2.8.9}$$

惟一地确定一个概率测度$P(\mathscr{B})$。反之，设$P(\mathscr{B})$是概率测度，那么存在惟一的分布函数$F(x)$使得(2.8.9)成立。

证明 先证定理的前半部分。令

$$\mathscr{A}=\{(a,b]\mid-\infty\leqslant a<b<+\infty\}\tag{2.8.10}$$

---

① 简称为柯氏概率空间。(2)、(3)、(4)段中的概率空间在概率论中已被广泛使用，且起着核心的作用。把它们命名为柯氏的则是作者的杜撰。

$$\mathscr{D}=\left\{\bigcup_{i=1}^{m}(a_i,b_i]\,|\,m\geqslant 1,(a_1,b_1],\cdots,(a_m,b_m]\text{互不相交}\right\}$$

(2.8.11)

那么 $\mathscr{B}=\sigma(\mathscr{A})=\sigma(\mathscr{D})$。容易验证 $\mathscr{D}$ 是事件域。

设 $F(x)$ 是给定的分布函数。那么用(2.8.9)可以定义 $P(\mathscr{A})$。对 $\mathscr{D}$ 中任意的事件 $\bigcup_{i=1}^{m}(a_i,b_i]$，令

$$P\left(\bigcup_{i=1}^{m}(a_i,b_i]\right)=\sum_{i=1}^{m}[F(b_i)-F(a_i)] \qquad (2.8.12)$$

可以验证 $P(\mathscr{D})$ 是单值集函数，并且是 $\mathscr{D}$ 上的概率测度[16]。现在，应用定理 2.7.4 即得前半部分的结论。

再证"反之"部分。设 $P(\mathscr{B})$ 是给定的概率测度。对任意的实数 x，定义

$$F(x)=P((-\infty,x]) \qquad (2.8.13)$$

由于 $(a,b]=(-\infty,b]\setminus(-\infty,a]$，应用概率的包含可减性推出(2.8.9)成立。

由(2.8.9)和概率的非负性得出 $F(x)$ 是非负的、单调不减的函数。注意到 $(-\infty,x]=\bigcap_{m=1}^{\infty}(-\infty,x+\varepsilon_m]$，这里 ε_m 非负，$\varepsilon_m\downarrow 0$。应用概率测度的上连续性(定理 2.7.2)得

$$F(x)=P\left(\bigcap_{m=1}^{\infty}(-\infty,x+\varepsilon_m]\right)=\lim_{m\to\infty}P((-\infty,x+\varepsilon_m])$$

$$=\lim_{m\to\infty}F(x+\varepsilon_m)$$

得证 $F(x)$ 满足右连续性条件。类似地可证 $F(x)$ 满足规范性条件(2.8.8)。于是证明了 $F(x)$ 是分布函数和"反之"的存在性部分。由于(2.8.9)含有(2.8.13)，因此 $F(x)$ 是惟一的。 **证毕**

应用勒贝格—斯蒂尔吉斯(T. Stieltjes)积分的知识容易得出

推论 波莱尔试验 $(\mathbf{R},\mathscr{B})$ 上的概率测度 $P(\mathscr{B})$ 和分布函数 $F(x)$ 之间用公式(2.8.13)和

$$P(A)=\int_A \mathrm{d}F(x) \quad (A\in\mathscr{B}) \tag{2.8.14}$$

相互惟一确定。这里的积分是勒贝格—斯蒂尔吉斯积分。

定义 2.8.6 称(2.8.13)规定的点函数 $F(x)$ 为概率测度 $P(\mathscr{B})$ 的分布函数。

2.8.3 柯尔莫哥洛夫概率空间：n 维情形

定义 2.8.7 称概率空间$(\mathbf{R}^n,\mathscr{B}^n,P)$为 n 维柯尔莫哥洛夫的，如果$(\mathbf{R}^n,\mathscr{B}^n)$是 n 维波莱尔试验。

定义 2.8.8 称定义在 n 维欧氏空间 $\mathbf{R}^n$ 上的非负实值函数 $F(x_1,x_2,\cdots,x_n)$，$(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$ 为 n 元分布函数，如果 $F(x_1,x_2,\cdots,x_n)$具有下列 4 个性质：

(1) 对于每个自变量 $x_i(i=1,2,\cdots,n)$单调不减；

(2) 对于每个自变量 $x_i(i=1,2,\cdots,n)$右连续；

(3) 对于任意的 $i(1\leqslant i\leqslant n)$，当 $x_i\to-\infty$时有

$$F(x_1,x_2,\cdots,x_n)\to 0 \tag{2.8.15}$$

当 $x_1\to+\infty,x_2\to+\infty,\cdots,x_n\to+\infty$时有

$$F(x_1,x_2,\cdots,x_n)\to 1 \tag{2.8.16}$$

(4) 对于任意的$-\infty<a_i<b_i<+\infty(i=1,2,\cdots,n)$有

$$F_0-\sum_{i=1}^{n}F_i+\sum_{1\leqslant i<j\leqslant n}F_{ij}-\sum_{1\leqslant i<j<k\leqslant n}F_{ijk}+\cdots+(-1)^nF_{12\cdots n}\geqslant 0 \tag{2.8.17}$$

其中 $F_0=F(b_1,b_2,\cdots,b_n)$，对于任意的 $1\leqslant i<j<\cdots<m\leqslant n$ 有

$$F_{ij\cdots m}=F(x_1,x_2,\cdots,x_n)\Bigg|_{\substack{x_r=a_r,r=i,j,\cdots,m\\ x_s=b_s,s\neq i,j,\cdots,m}} \tag{2.8.18}$$

定理 2.8.3 设$(\mathbf{R}^n,\mathscr{B}^n)$是 n 维波莱尔试验。那么 n 元分布函数 $F(x_1,x_2,\cdots,x_n)$用公式

$$P\Big(\prod_{i=1}^{n}(a_i,b_i]\Big)=F_0-\sum_{i=1}^{n}F_i+\sum_{1\leqslant i<j\leqslant n}F_{ij}-\cdots+(-1)^n F_{12\cdots n}$$

(2.8.19)

惟一地确定一个概率测度 $P(\mathscr{B}^n)$。反之，设 $P(\mathscr{B}^n)$ 是概率测度，那么存在惟一的 n 元分布函数 $F(x_1,x_2,\cdots,x_n)$ 使得(2.8.19)成立。

推论 n 维波莱尔试验 $(\mathbf{R}^n,\mathscr{B}^n)$ 上的概率测度 $P(\mathscr{B}^n)$ 和 n 元分布函数 $F(x_1,x_2,\cdots,x_n)$ 之间用公式

$$F(x_1,x_2,\cdots,x_n)=P\Big(\prod_{i=1}^{n}(-\infty,x_i]\Big) \qquad (2.8.20)$$

$$P(A)=\int_A \mathrm{d}_{x_1}\mathrm{d}_{x_2}\cdots\mathrm{d}_{x_n}F(x_1,x_2,\cdots,x_n) \quad (A\in\mathscr{B}^n)$$

(2.8.21)

相互惟一确定。这里的积分是 n 维勒贝格—斯蒂尔吉斯积分。

定理 2.8.2 和 2.8.3 是勒贝格—斯蒂尔吉斯测度的经典定理。n 维情形的证明与一维情形类似[22]，从略。

定义 2.8.9 称(2.8.20)规定的点函数 $F(x_1,x_2,\cdots,x_n)$ 为概率测度 $P(\mathscr{B}^n)$ 的 n 元分布函数。

2.8.4 柯尔莫哥洛夫概率空间：无限维情形

用 T 表示任意的无限指标集。

定义 2.8.10 称概率空间 $(\mathbf{R}^T,\mathscr{B}^T,P)$ 为 T 维柯尔莫哥洛夫的，如果 $(\mathbf{R}^T,\mathscr{B}^T)$ 是 T 维波莱尔试验。

定义 2.8.11 对固定的 $t_1,t_2,\cdots,t_n(\in T)$，引进 n 元分布函数 $F(t_1,x_1;t_2,x_2;\cdots;t_n,x_n)$。称

$$F=\{F(t_1,x_1;t_2,x_2;\cdots;t_n,x_n)\,|\,n\geqslant 1,t_1,t_2,\cdots,t_n\in T\}$$

(2.8.22)

为 $(\mathbf{R}^T,\mathscr{B}^T)$ 上的有穷维分布函数族。

定义 2.8.12 设 F 是 $(\mathbf{R}^T,\mathscr{B}^T)$ 上的有穷维分布函数族。称 F

为相容的，如果 F 满足下列两个条件：

(1) 设 $(\alpha_1,\alpha_2,\cdots,\alpha_n)$ 是 $(1,2,\cdots,n)$ 的任意的一种排列，则

$$F(t_{\alpha_1},x_{\alpha_1};t_{\alpha_2},x_{\alpha_2};\cdots;t_{\alpha_n},x_{\alpha_n}) = F(t_1,x_1;t_2,x_2;\cdots;t_n,x_n) \tag{2.8.23}$$

(2) 如果 $m<n$，则

$$F(t_1,x_1;t_2,x_2;\cdots;t_m,x_m) = \lim_{x_{m+1},\cdots,x_n\to+\infty} F(t_1,x_1;t_2,x_2;\cdots;t_n,x_n) \tag{2.8.24}$$

定理 2.8.4 设 $(\mathbf{R}^T,\mathscr{B}^T)$ 是 T 维波莱尔试验。那么 $(\mathbf{R}^T,\mathscr{B}^T)$ 上相容的有穷维分布函数族 F 用公式

$$P\Big(\prod_{i=1}^{n}(a_{t_i},b_{t_i}]\times\prod\mathbf{R}\Big)=F_0-\sum_{i=1}^{n}F_i+\sum_{1\leqslant i<j\leqslant n}F_{ij}-\cdots+(-1)^nF_{12\cdots n} \tag{2.8.25}$$

惟一地确定一个概率测度 $P(\mathscr{B}^T)$。其中左方的事件是(2.6.37)规定的柱事件，右方是(2.8.17)规定的数值，但要用 $F(t_1,x_1;t_2,x_2;\cdots;t_n,x_n)$ 代替 $F(x_1,x_2,\cdots,x_n)$，且用 a_{t_i}、b_{t_i} 分别代替 a_i、b_i $(1\leqslant i\leqslant n)$。

反之，设 $P(\mathscr{B}^T)$ 是概率测度。对任意的实数 $x_1,x_2,\cdots,x_n$ 和任意的 $t_1,t_2,\cdots,t_n\in T$，令

$$F(t_1,x_1;t_2,x_2;\cdots;t_n,x_n)=P\Big(\prod_{i=1}^{n}(-\infty,x_i]\times\prod\mathbf{R}\Big) \tag{2.8.26}$$

$$F=\{F(t_1,x_1;t_2,x_2;\cdots;t_n,x_n)\mid n\geqslant 1,t_1,t_2,\cdots,t_n\in T\} \tag{2.8.27}$$

那么 F 是相容的有穷维分布函数族，并且使得(2.8.25)成立。

这是著名的柯尔莫哥洛夫存在定理[1]。它是随机过程理论的基础。

定义 2.8.13 称(2.8.27)规定的 F 为概率测度 $P(\mathscr{B}^T)$ 的有穷维分布函数族。

2.8.5 注记

测度论中众所周知的一个结论是，如果 n 元分布函数不是离散型的，那么用它确定的概率测度 $P(\mathscr{B}^n)$ 不能扩张到 $(\mathbf{R}^n, \mathscr{F}_{11})$（2.4 节例 11）上。这是概率论很少使用随机试验 $(\mathbf{R}^n, \mathscr{F}_{11})$ 的原因。

2.9 第Ⅴ组公理：条件概率公理组

在随机试验 $(\Omega, \mathscr{F})$ 上可以赋予许多的概率测度。研究这些概率测度之间的关系是概率论的一项重要任务。第Ⅴ组公理探讨条件 $\mathscr{C}$ 与概率测度 $P(\mathscr{F})$ 之间的内在联系，抽象出条件概率和独立性的概念。

随机试验 $(\Omega_1, \mathscr{F}_1)$ 和 $(\Omega_2, \mathscr{F}_2)$ 是联合试验 $(\Omega_1 \circ \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2)$ 的子试验。因此，条件概率和独立性也是研究概率空间之间内在联系的重要工具。

2.9.1 直观背景

在本节中假定 $(\Omega, \mathscr{F}, P)$ 是用 2.7 节中与建模有关的式子

$$\text{实验 } \mathscr{E}(\text{在划定条件 } \mathscr{C} \text{ 下}) \Longrightarrow (\Omega, \mathscr{F}, P)$$

产生的概率空间。让我们考察条件 $\mathscr{C}$ 变化时产生的概率空间是如何变化的。

(1) 假定 $B \in \mathscr{F}, P(B) > 0$。如果把“事件 B 已经出现”添加到条件 $\mathscr{C}$ 上，得到的新条件用 $\mathscr{C}^*$ 表示。

根据常识，在条件 $\mathscr{C}^*$ 下可把芽事件选为 B 的所有子事件，即

$$\mathscr{F}_B = \{A \mid A \in \mathscr{F}, A \subset B\} \qquad (2.9.1)$$

因此条件 $\mathscr{C}^*$ 产生随机试验 $(\Omega, \sigma(\mathscr{F}_B))$。容易看出 $\overline{B}$ 是 $\sigma(\mathscr{F}_B)$ 中的原子事件。代替(2.7.4)我们有

$$\text{实验 } \mathscr{E}(\text{在划定条件 } \mathscr{C}^* \text{ 下}) \Longrightarrow (\Omega, \sigma(\mathscr{F}_B), P^*) \tag{2.9.2}$$

其中 P^* 是待探讨的概率测度。显然条件 $\mathscr{C}^*$ 蕴含

$$P^*(B) = 1, P^*(\overline{B}) = 0 \tag{2.9.3}$$

首先讨论事件族 $\mathscr{F}_B$。经验告诉我们,当条件 $\mathscr{C}$ 加强为 $\mathscr{C}^*$ 时族中事件出现的可能性都将增大,即 $P^*(A) \geqslant P(A)(A \in \mathscr{F}_B)$。增大的幅度是多少?合理的假定是它们按比例地增大。即成立

$$\frac{P^*(A)}{P(A)} = \frac{P^*(A_1)}{P(A_1)} \quad (A, A_1 \in \mathscr{F}_B, P(A) > 0, P(A_1) > 0) \tag{2.9.4}$$

特别,取 $A_1 = B$ 时成为

$$P^*(A) = \frac{P(A)}{P(B)} \quad (A \in \mathscr{F}_B) \tag{2.9.5}$$

然后讨论事件族 $\sigma(\mathscr{F}_B)$。对任意的 $A \in \sigma(\mathscr{F}_B)$,令 $D = A \cap B$,则 $D \in \mathscr{F}_B$。由 $\overline{B}$ 的原子性推出 $A = D$ 和 $A = D \cup \overline{B}$ 中有且只有一个成立。不论何种情况,令 $P^*(A) = P^*(D)$ 是合理的。于是(2.9.5)被扩张成

$$P^*(A) = \frac{P(AB)}{P(B)} \quad (A \in \sigma(\mathscr{F}_B)) \tag{2.9.6}$$

并且实现了(2.9.2)反映的因果关系。

(2) 采用类似于第 2.7 节末注记 1 的增补方法。在 $(\Omega, \sigma(\mathscr{F}_B), P^*)$ 中已知 $P(\overline{B}) = 0$,令

$$\mathscr{F}_{\overline{B}} = \{A \mid A \in \mathscr{F}, A \subset \overline{B}\} \tag{2.9.7}$$

$$P^*(A) = 0 \quad (A \in \mathscr{F}_{\overline{B}}) \tag{2.9.8}$$

显然 $\mathscr{F} = \sigma(\mathscr{F}_B \cup \mathscr{F}_{\overline{B}})$,并且对任意的 $A \in \mathscr{F}$ 有

$$A = (AB) \cup (A\overline{B})$$

因此,规定

$$P^*(A) = P^*(AB) + P^*(A\overline{B}) = \frac{P(AB)}{P(B)} \quad (A \in \mathscr{F}) \tag{2.9.9}$$

是合理的。于是我们又把 $P^*(\sigma(\mathscr{F}_B))$扩张成 $P^*(\mathscr{F})$。容易验证$(\Omega,\sigma(\mathscr{F}_B),P^*)$是$(\Omega,\mathscr{F},P^*)$的同因子概率空间。(2.9.2)可以代之为

$$\text{实验 } \mathscr{E}(\text{在划定条件 } \mathscr{C}^* \text{ 下}) \Longrightarrow (\Omega,\mathscr{F},P^*) \tag{2.9.10}$$

(3) 把 $P^*(A),A\in\mathscr{F}$ 改记为 $P(A|B),A\in\mathscr{F}$;把 $P^*(\mathscr{F})$改记为 $P(\mathscr{F}|B)$。比较(2.7.4)和(2.9.10)后自然称 $P(\mathscr{F}|B)$为在事件 B 出现的条件下的概率测度(注意,条件 $\mathscr{C}$ 仍然保持),简称为关于 B 的条件概率(或条件概率测度)。称数值 $P(A|B)$为事件 A 关于 B 的条件概率值。在新记号下(2.9.9)成为众所周知的等式

$$P(A|B)=\frac{P(AB)}{P(B)}\quad(A\in\mathscr{F}) \tag{2.9.11}$$

(4) 概率的 4 种解释都满足(2.9.11):

① 在概率的古典定义中由(2.7.8)得出

$$P(A|B)=\frac{AB\text{ 含样本点的个数}}{B\text{ 含样本点的个数}} \tag{2.9.12}$$

它表明,当条件 $\mathscr{C}$ 加强成 $\mathscr{C}^*$ 时 B 起着必然事件 Ω 的作用,古典定义的赋概方法继续保持。

② 在几何概率中由(2.7.11)得出

$$P(A|B)=\frac{\mu(AB)}{\mu(B)} \tag{2.9.13}$$

它表明,当条件 $\mathscr{C}$ 加强成 $\mathscr{C}^*$ 时 B 起着必然事件 D 的作用,几何概率的赋概方法继续保持。

③ 在概率的频率解释中由(1.3.15)得出

$$P(A|B)=\lim_{\nu_n(B)\to\infty}\frac{\nu_n(AB)}{\nu_n(B)} \tag{2.9.14}$$

这里使用 $\nu_n(B)\to\infty$是合理的,因为 $P(B)=\lim\limits_{n\to\infty}\frac{\nu_n(B)}{n}>0$。它表明,从理想的无限序列中删除事件 B 不出现的实验,剩下的实验

仍然组成一个理想的无限序列。这时 B 起着必然事件的作用，频率解释的赋概方法继续有效。

④ 在概率的主观解释中条件概率像概率一样被直接定义，(2.9.11)是相容性要求的后果[2,8]。

2.9.2 条件概率公理组

引理 2.9.1 设$(\Omega,\mathscr{F},P)$是概率空间。令

$$\mathscr{F}^* = \{B \mid B\in\mathscr{F}, P(B)>0\}^{①} \tag{2.9.15}$$

那么二元集函数

$$P(A\mid B) \overset{\text{def}}{=\!=} \frac{P(AB)}{P(B)}, A\in\mathscr{F}, B\in\mathscr{F}^* \tag{2.9.16}$$

具有下列 3 个性质：

(1) $0\leqslant P(A\mid B)\leqslant 1$； (2.9.17)

(2) $P(\Omega\mid B)=1$； (2.9.18)

(3) 对任意的 $B\in\mathscr{F}^*$ 和 $\mathscr{F}$ 中有限或可列个互不相容的事件 $A_i(i\in I)$成立

$$P\Big(\bigcup_{i\in I} A_i \mid B\Big) = \sum_{i\in I} P(A_i\mid B) \tag{2.9.19}$$

证明 (1)和(2)显然。(3)由概率 $P(\mathscr{F})$遵守公理 IV_3 得出。

证毕

推论 对固定的 $B\in\mathscr{F}^*$，一元集函数 $P(\mathscr{F}\mid B)$是$(\Omega,\mathscr{F})$上的概率测度。二元集函数 $P(\mathscr{F}\mid\mathscr{F}^*)$是$(\Omega,\mathscr{F})$上的概率测度族 $\{P(\mathscr{F}\mid B)\mid B\in\mathscr{F}^*\}^{②}$。

众所周知，定义函数的两个要素是对应律和定义域。即使对应律相同，不同的定义域也表示不同的函数。为了强调定义域的重要

---

① 今后，事件 σ 域的右上标"*"总是起这种作用。

② $P(\mathscr{F}\mid B)$是 $P(A\mid B)$, $A\in\mathscr{F}$ 的简写；$P(\mathscr{F}\mid\mathscr{F}^*)$是 $P(A\mid B)$, $A\in\mathscr{F}$, $B\in\mathscr{F}^*$的简写。参看定义 2.7.1 下的脚注。

性，我们使用定义 2.7.1 下的脚注，且在上面的推论中使用记号 $P(\mathscr{F}|B)$和 $P(\mathscr{F}|\mathscr{F}^*)$。为了书写方便，今后把 $P(\mathscr{F}|\mathscr{F}^*)$简记为 $P(\mathscr{F}|\mathscr{F})$。即是说，$\mathscr{F}$ 是否有右上标“ * ”，请读者自己判断。

设 $y=f(x), x\in\mathscr{D}$ 是任意函数。用它获得新函数 $y=f(x)$，$x\in\mathscr{D}\cap\mathscr{D}_1$的演算称为 $f(\mathscr{D})$在 $\mathscr{D}_1$ 上的限制，记为 $f(\mathscr{D})\upharpoonright\mathscr{D}_1$。即

$$f(\mathscr{D}\cap\mathscr{D}_1)=f(\mathscr{D})\upharpoonright\mathscr{D}_1 \tag{2.9.20}$$

特别，当 $\mathscr{D}_1\subset\mathscr{D}$ 时成为 $f(\mathscr{D}_1)=f(\mathscr{D})\upharpoonright\mathscr{D}_1$。

假定 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 是 $\mathscr{F}$ 的两个子事件 σ 域。应用限制演算，用 $P(\mathscr{F}|\mathscr{F})$可以得到如下 3 类集函数：

$$P(\mathscr{F}_1|\mathscr{F}_2)，\text{即 } P(A|B), A\in\mathscr{F}_1, B\in\mathscr{F}_2^* \tag{2.9.21}$$

$$P(A|\mathscr{F}_2)，\text{即 } P(A|B), B\in\mathscr{F}_2^* \tag{2.9.22}$$

$$P(\mathscr{F}_1|B)，\text{即 } P(A|B), A\in\mathscr{F}_1 \tag{2.9.23}$$

第 V 组公理用这 3 类集函数揭示条件与概率之间的内在联系。在这组公理中$(\Omega,\mathscr{F},P)$是用(2.7.4)产生的概率空间。

公理 V_1(事件条件公理)　设 $B\in\mathscr{F}^*$。把“事件 B 已经出现”添加到条件 $\mathscr{C}$ 上得到新条件 $\mathscr{C}^*$。那么条件 $\mathscr{C}^*$ 在子试验$(\Omega,\mathscr{F}_1)$上赋予概率测度 $P(\mathscr{F}_1|B)$。

公理 V_2(子试验条件公理)　把“子试验$(\Omega,\mathscr{F}_2)$进行”添加到条件 $\mathscr{C}$ 上得到新条件 $\mathscr{C}^{**}$。那么条件 $\mathscr{C}^{**}$ 在子试验$(\Omega,\mathscr{F}_1)$上赋予概率测度族 $P(\mathscr{F}_1|\mathscr{F}_2)$。

定义 2.9.1　给定概率空间$(\Omega,\mathscr{F},P)$。假定 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 是 $\mathscr{F}$ 的子事件 σ 域，$A\in\mathscr{F}$，$B\in\mathscr{F}^*$，那么

(1) 称 $P(\mathscr{F}_1|\mathscr{F}_2)$为子试验$(\Omega,\mathscr{F}_1)$关于$(\Omega,\mathscr{F}_2)$的条件概率测度族，简称为 $\mathscr{F}_1$ 关于 $\mathscr{F}_2$ 的条件概率。

(2) 称 $P(A|\mathscr{F}_2)$为事件 A 关于子试验$(\Omega,\mathscr{F}_2)$的条件概率值族，简称为 A 关于 $\mathscr{F}_2$ 的条件概率。

(3) 称 $P(\mathscr{F}_1|B)$为子试验$(\Omega,\mathscr{F}_1)$关于事件 B 的条件概率测度,简称为 $\mathscr{F}_1$ 关于 B 的条件概率。

(4) 称 $P(A|B)$为事件 A 关于 B 的条件概率值,简称为 B 关于 A 的条件概率。

现今概率论尚未引进前两类条件概率,第 3 类虽然偶尔使用,但也没有正式定义。由于笼统地使用一个个的条件概率值 $P(A|B)$,使得条件概率成为一个直观性含糊的概念。

概率论的深入发展必然经常涉及前两类条件概率。于是,现代概率论借助随机变量的概念,应用测度论的一个抽象定理引进两类点函数,并用它们代替前两类条件概率。由于这两类点函数在很一般的条件下惟一地确定这两类条件概率,因此"代替"取得了巨大的成功。可是,由于不适当地把这两类点函数也称为条件概率,使得条件概率成为难理解的概念。

点函数形式的"条件概率"已成为现代概率论中最基本的概念之一,也是进行概率计算的最重要的工具之一,本组公理为它们提供了直观的、形象的概率论解释(见 2.10 节)。

2.9.3 两个子试验的相互独立性

可能是独立性的直观背景和重要作用,人们赞誉"独立性是一个使概率论区别于测度论的特有概念"[①]。可是,独立性的原有定义却是生硬的,人为造作的,与它的直观性和形象性不相称。第Ⅴ组公理的引入有望改变这种不合理的现象。

条件概率 $P(\mathscr{F}_1|\mathscr{F}_2)$ 和 $P(\mathscr{F}_2|\mathscr{F}_1)$ 反映子概率空间 $(\Omega,\mathscr{F}_1,P(\mathscr{F}_1))$和$(\Omega,\mathscr{F}_2,P(\mathscr{F}_2))$之间的联系。最简单的联系是"无联系",即 $\mathscr{F}_1$ 中事件的出现与子试验$(\Omega,\mathscr{F}_1)$是否进行无关;同样,$\mathscr{F}_2$ 中事件的出现与$(\Omega,\mathscr{F}_2)$是否进行无关。应用公理

---

① 《中国大百科全书·数学》,第 243 页。

V₂,“无联系”可以用数学式

$$P(\mathscr{F}_1|\mathscr{F}_2)=P(\mathscr{F}_1);P(\mathscr{F}_2|\mathscr{F}_1)=P(\mathscr{F}_2) \quad (2.9.24)$$

表达,于是产生独立性概念。

定义 2.9.2 称子试验$(\Omega,\mathscr{F}_1)$和$(\Omega,\mathscr{F}_2)$关于概率测度$P(\mathscr{F})$相互独立,简称为$\mathscr{F}_1$和$\mathscr{F}_2$相互独立,如果成立$P(\mathscr{F}_1|\mathscr{F}_2)=P(\mathscr{F}_1)$,即

$$P(A|B)=P(A) \quad (A\in\mathscr{F}_1,B\in\mathscr{F}_2^*) \quad (2.9.25)$$

定义 2.9.3 称事件A和B关于概率测度$P(\mathscr{F})$相互独立,简称为A和B独立,如果存在相互独立的子试验$(\Omega,\mathscr{F}_1)$和$(\Omega,\mathscr{F}_2)$使得$A\in\mathscr{F}_1,B\in\mathscr{F}_2$。

由两个定义直接推出

引理 2.9.2 (1) 退化子试验$(\Omega_0,\mathscr{F}_0)$与任何子试验皆相互独立。

(2) $\varnothing$和Ω与任何事件皆相互独立。

(3) 如果A和B独立,那么三对事件$\bar{A}$和B,A和$\bar{B}$,$\bar{A}$和$\bar{B}$也独立。

定理 2.9.1 设$(\Omega,\mathscr{F},P)$是概率空间,$\mathscr{F}_1$和$\mathscr{F}_2$是$\mathscr{F}$的两个子事件σ域。那么$\mathscr{F}_1$和$\mathscr{F}_2$相互独立的必要和充分条件是

$$P(AB)=P(A)P(B) \quad (A\in\mathscr{F}_1,B\in\mathscr{F}_2) \quad (2.9.26)$$

证明 必要性:当$P(B)=0$时(2.9.26)显然成立。下设$P(B)>0$。由(2.9.16)得出$P(AB)=P(A|B)P(B)$。应用(2.9.25)即得(2.9.26)。

充分性:当$B\in\mathscr{F}_2^*$时(2.9.26)含有$P(A)=\dfrac{P(AB)}{P(B)}$。应用(2.9.16)即得(2.9.25)。 **证毕**

推论 1 $P(\mathscr{F}_1|\mathscr{F}_2)=P(\mathscr{F}_1)$当且仅当$P(\mathscr{F}_2|\mathscr{F}_1)=P(\mathscr{F}_2)$。

推论 2(0-1 律) 设$\mathscr{F}_1$和$\mathscr{F}_2$相互独立,那么

$$\mathscr{F}_1 \cap \mathscr{F}_2 \subset \{A | P(A) = 0 \text{ 或 } 1\} \tag{2.9.27}$$

推论 3(0-1 律) 如果$\mathscr{F}_1$和自己相互独立,那么

$$\mathscr{F}_1 \subset \{A | P(A) = 0 \text{ 或 } 1\} \tag{2.9.28}$$

推论 4 独立乘积概率空间$(\Omega_1 \times \Omega_2, \mathscr{F}_1 \times \mathscr{F}_2, P_1 \times P_2)$中子试验$(\Omega_1 \times \Omega_2, \mathscr{F}_1)$和$(\Omega_1 \times \Omega_2, \mathscr{F}_2)$相互独立。

推论 5 给定概率空间$(\Omega, \mathscr{F}, P)$,假定子事件σ域$\mathscr{F}_1$和$\mathscr{F}_2$相互独立。那么$\mathscr{F}_1$和$\mathscr{F}_2$在子概率空间$(\Omega, \mathscr{F}_1 \otimes \mathscr{F}_2, P)$中也相互独立。

于是,只要$(\Omega, \mathscr{F}, P)$包含两个独立的事件,那么它包含两个独立的子试验(不妨设$(\Omega, \mathscr{F}_1)$和$(\Omega, \mathscr{F}_2)$),从而包含子概率空间$(\Omega, \mathscr{F}_1 \otimes \mathscr{F}_2, P)$。

定理 2.9.2 设$(\Omega, \mathscr{F}, P)$是概率空间,$\mathscr{F}_1$与$\mathscr{F}_2$是相互独立的子事件σ域。那么子概率空间$(\Omega, \mathscr{F}_1 \otimes \mathscr{F}_2, P)$具有性质:假定$A \in \mathscr{F}_1, B \in \mathscr{F}_2$,则

(1) (2.9.26)成立。由此推出概率测度$P(\mathscr{F}_1 \otimes \mathscr{F}_2)$由$P(\mathscr{F}_1)$和$P(\mathscr{F}_2)$惟一确定;

(2) $P(A \cap B) = 0$当且仅当$P(A) = 0, P(B) = 0$中至少有一个成立;

(3) $\mathscr{F}_1 \cap \mathscr{F}_2 \subset \{C | P(C) = 0 \text{ 或 } 1\}$;

(4) 如果$P(A)$和$P(B)$皆不等于0或1,那么

$$A \cap B \overline{\in} \mathscr{F}_1, A \cap B \overline{\in} \mathscr{F}_2$$

证明 (1) (2.9.26)由定理 2.9.1 得出。注意到$\mathscr{F}_1 \otimes \mathscr{F}_2$的芽事件集$\mathscr{F}_1 * \mathscr{F}_2$是由(2.6.1)规定,因此$P(\mathscr{F}_1)$和$P(\mathscr{F}_2)$惟一地确定$\mathscr{F}_1 * \mathscr{F}_2$中事件的概率。应用定理 2.7.5 得出$P(\mathscr{F}_1)$和$P(\mathscr{F}_2)$惟一地确定$P(\mathscr{F}_1 \otimes \mathscr{F}_2)$。

(2) 是(2.9.26)的直接后果。(3)是 0-1 律。下证(4)。若不能,设$A \cap B \in \mathscr{F}_1$。应用(2.9.26)得出

$$P(A \cap B) = P([A \cap B] \cap B) = P(A \cap B)P(B)$$

与 $P(B) \neq 0$ 或 1 矛盾，得证 $A \cap B \overline{\in} \mathscr{F}_1$。同理 $A \cap B \overline{\in} \mathscr{F}_2$。

证毕

把这个定理与引理 2.6.2、定理 2.7.6（$n=2$ 的情形）比较得出，$(\Omega, \mathscr{F}_1 \otimes \mathscr{F}_2, P)$ 有点像独立乘积概率空间的推广，那里的精确等式与包含式现在推广成不计零概率集意义下的"等式"与"包含式"。因此，直观上可以认为 $(\Omega, \mathscr{F}_1 \otimes \mathscr{F}_2, P)$ 本质上类似于独立乘积概率空间。

定理 2.9.3 设 $(\Omega, \mathscr{F}, P)$ 是概率空间，$A, B \in \mathscr{F}$。那么事件 A 和 B 独立当且仅当

$$P(AB) = P(A)P(B) \tag{2.9.29}$$

证明 必要性由定理 2.9.1 推出。下证充分性。设 A 和 B 满足(2.9.29)。应用概率的包含可减性得

$$\begin{aligned} P(\overline{A}B) &= P(B) - P(AB) = [1 - P(A)]P(B) \\ &= P(\overline{A})P(B) \end{aligned} \tag{2.9.30}$$

由 A 与 B 的对称性得 $P(A\overline{B}) = P(A)P(\overline{B})$，把(2.9.30)中 B 换为 $\overline{B}$ 得 $P(\overline{A}\overline{B}) = P(\overline{A})P(\overline{B})$。现在令

$$\mathscr{F}_1 = \{\varnothing, A, \overline{A}, \Omega\}; \mathscr{F}_2 = \{\varnothing, B, \overline{B}, \Omega\} \tag{2.9.31}$$

由定理 2.9.1 推出 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 相互独立。得证 A 和 B 独立。

证毕

2.9.4 n 个子试验的相互独立性

用 $(n_1, n_2, \cdots, n_n)$ 表示 $(1, 2, \cdots, n)$ 的任意一种排列。

定义 2.9.4 给定概率空间 $(\Omega, \mathscr{F}, P)$，设 $(\Omega, \mathscr{F}_i)$ $(i = 1, 2, \cdots, n)$ 是它的 n 个子试验。如果对任意的 k $(1 \leqslant k \leqslant n-1)$，子试验 $(\Omega, \mathscr{F}_{n_1} \otimes \mathscr{F}_{n_2} \otimes \cdots \otimes \mathscr{F}_{n_k})$ 和 $(\Omega, \mathscr{F}_{n_{k+1}} \otimes \mathscr{F}_{n_{k+2}} \otimes \cdots \otimes \mathscr{F}_{n_n})$ 相互独立，则称 n 个子试验 $(\Omega, \mathscr{F}_1), (\Omega, \mathscr{F}_2), \cdots, (\Omega, \mathscr{F}_n)$ 相互独立，简称为 n 个子事件 σ 域 $\mathscr{F}_1, \mathscr{F}_2, \cdots, \mathscr{F}_n$ 相互独立。

• **定义 2.9.5** 设 $A_i(i=1,2,\cdots,n)$是$(\Omega,\mathscr{F},P)$中的 n 个事件，如果存在 n 个相互独立的子事件 σ 域 $\mathscr{F}_1,\mathscr{F}_2,\cdots,\mathscr{F}_n$ 使得 $A_i\in\mathscr{F}_i(i=1,2,\cdots,n)$，则称 n 个事件 $A_1,A_2,\cdots,A_n$ 相互独立。

引理 2.9.3 设事件 $A_1,A_2,\cdots,A_n$ 相互独立。那么对任意的 k，事件 $A_{n_1},A_{n_2},\cdots,A_{n_k},\overline{A}_{n_{k+1}},\cdots,\overline{A}_{n_n}$ 也相互独立。

证明 它是定义 2.9.5 的直接推论。 **证毕**

引理 2.9.4 设$(\Omega,\mathscr{F},P)$是概率空间，$B\in\mathscr{F}$。用 $\mathscr{D}_B$ 表示与 B 相互独立的事件集合，即

$$\mathscr{D}_B=\{A\,|\,A\in\mathscr{F},P(AB)=P(A)P(B)\} \quad (2.9.32)$$

那么 $\mathscr{D}_B$ 是 λ 系，即它具有下列 4 个性质：

(1) $\varnothing,\Omega\in\mathscr{D}_B$；

(2) $A_1,A_2\in\mathscr{D}_B$，且 $A_1\cap A_2=\varnothing$，则 $A_1\cup A_2\in\mathscr{D}_B$；

(3) $A_1,A_2\in\mathscr{D}_B$，且 $A_1\supset A_2$，则 $A_1\backslash A_2\in\mathscr{D}_B$；

(4) $A_i\in\mathscr{D}_B(i\in I)$，且 $A_1\subset A_2\subset\cdots\subset A_i\subset\cdots$，则 $\bigcup\limits_{i=1}^{\infty}A_i\in\mathscr{D}_B$。

证明 (1) 显然。

(2) 由$(A_1B)\cap(A_2B)=\varnothing$和 $A_1,A_2\in\mathscr{D}_B$ 得出

$$\begin{aligned}&P([A_1\cup A_2]\cap B)\\&=P(A_1B)+P(A_2B)\\&=[P(A_1)+P(A_2)]P(B)\\&=P(A_1\cup A_2)P(B)\end{aligned}$$

故 $A_1\cup A_2\in\mathscr{D}_B$。

(3) 注意到 $A_1B\supset A_2B$，由概率的包含可减性得出

$$\begin{aligned}&P([A_1\backslash A_2]\cap B)\\&=P(A_1B)-P(A_2B)\\&=[P(A_1)-P(A_2)]P(B)\\&=P(A_1\backslash A_2)P(B)\end{aligned}$$

故 $A_1\backslash A_2\in\mathscr{D}_B$。

(4) 诸 $A_i\backslash A_{i-1}(i=1,2,3,\cdots,A_0=\varnothing)$ 互不相交，并且 $\bigcup_{i=1}^{\infty}A_i=\bigcup_{i=1}^{\infty}(A_i\backslash A_{i-1})$。应用概率的 σ 可加性和已证的(3)得

$$
\begin{aligned}
P\left(\left[\bigcup_{i=1}^{\infty}A_i\right]\cap B\right)&=\sum_{i=1}^{\infty}P([A_i\backslash A_{i-1}]\cap B)\\
&=\sum_{i=1}^{\infty}P(A_i\backslash A_{i-1})P(B)\\
&=P\left(\bigcup_{i=1}^{\infty}A_i\right)P(B)
\end{aligned}
$$

故 $\bigcup_{i=1}^{\infty}A_i\in\mathscr{D}_B$。 **证毕**

定理 2.9.4 设 $(\Omega,\mathscr{F},P)$ 是概率空间，$(\Omega,\mathscr{F}_i)(i=1,2,\cdots,n)$ 是 n 个子试验。那么 $(\Omega,\mathscr{F}_1),(\Omega,\mathscr{F}_2),\cdots,(\Omega_n,\mathscr{F}_n)$ 相互独立的充分必要条件是对任意的 $A_i\in\mathscr{F}_i(i=1,2,\cdots,n)$ 成立

$$
P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n) \tag{2.9.33}
$$

证明 必要性：假定 $A_i\in\mathscr{F}_i(i=1,2,\cdots,n)$，则 $A_2A_3\cdots A_n\in\mathscr{F}_2\otimes\mathscr{F}_3\otimes\cdots\otimes\mathscr{F}_n$。由于 $\mathscr{F}_1$ 和 $\mathscr{F}_2\otimes\mathscr{F}_3\otimes\cdots\otimes\mathscr{F}_n$ 相互独立，由定理 2.9.1 得出

$$
P(A_1A_2\cdots A_n)=P(A_1)P(A_2A_3\cdots A_n)
$$

类似地可证 $P(A_2A_3\cdots A_n)=P(\Omega A_2)P(A_3\cdots A_n)$。重复这样的操作 $n-1$ 次即得(2.9.33)。

充分性：假定(2.9.33)成立。往证对任意的 k，$\mathscr{F}_{n_1}\otimes\mathscr{F}_{n_2}\otimes\cdots\otimes\mathscr{F}_{n_k}$ 和 $\mathscr{F}_{n_{k+1}}\otimes\mathscr{F}_{n_{k+2}}\otimes\cdots\otimes\mathscr{F}_{n_n}$ 相互独立。由定理 2.9.1 知，只须证对任意的

$$
\left.\begin{aligned}
E&\in\mathscr{F}_{n_1}\otimes\mathscr{F}_{n_2}\otimes\cdots\otimes\mathscr{F}_{n_k}\\
F&\in\mathscr{F}_{n_{k+1}}\otimes\mathscr{F}_{n_{k+2}}\otimes\cdots\otimes\mathscr{F}_{n_n}
\end{aligned}\right\} \tag{2.9.34}
$$

成立 $P(EF)=P(E)P(F)$。

选定 $\mathscr{F}_{n_1} * \mathscr{F}_{n_2} * \cdots * \mathscr{F}_{n_k}$ 中柱事件 $C=A_{n_1}A_{n_2}\cdots A_{n_k}$。由引理 2.9.4 得出，事件集合

$$\mathscr{D}_C = \{A \mid A \in \mathscr{F}, P(AC) = P(A)P(C)\}$$

是 λ 系。另一方面，(2.9.33)蕴含 $\mathscr{D}_C \supset \mathscr{F}_{n_{k+1}} * \mathscr{F}_{n_{k+2}} * \cdots * \mathscr{F}_{n_n}$，引理 2.6.1 保证包含式右方是 π 系。应用 λ 系方法得出[23,20]

$$\mathscr{D}_C \supset \mathscr{F}_{n_{k+1}} \otimes \mathscr{F}_{n_{k+2}} \otimes \cdots \otimes \mathscr{F}_{n_n} \tag{2.9.35}$$

现在，对(2.9.34)中的 F 引进

$$\mathscr{D}_F = \{A \mid A \in \mathscr{F}, P(AF) = P(A)P(F)\}$$

由引理 2.9.4 知 $\mathscr{D}_F$ 是 λ 系。由(2.9.35)得出 $C \in \mathscr{D}_F$。注意到 C 可以任意地选取，故 $\mathscr{D}_F \supset \mathscr{F}_{n_1} * \mathscr{F}_{n_2} * \cdots * \mathscr{F}_{n_k}$。由于包含式右方是 π 系，再次应用 λ 系方法得出

$$\mathscr{D}_F \supset \mathscr{F}_{n_1} \otimes \mathscr{F}_{n_2} \otimes \cdots \otimes \mathscr{F}_{n_k}$$

由此推出对(2.9.34)中的 E 和 F 成立 $P(EF)=P(E)P(F)$。充分性得证。　　**证毕**

推论 1　假定子事件 σ 域 $\mathscr{F}_1, \mathscr{F}_2, \cdots, \mathscr{F}_n$ 相互独立，那么其中任意 $m(m<n)$ 个子事件 σ 域也相互独立。

推论 2　独立乘积概率空间 $\left(\prod_{i=1}^{n} \Omega_i, \prod_{i=1}^{n} \mathscr{F}_i, \prod_{i=1}^{n} P_i\right)$ 中 n 个子试验 $\left(\prod_{i=1}^{n} \Omega_i, \mathscr{F}_i\right)$ $(i=1,2,\cdots,n)$ 相互独立。

推论 3　给定概率空间 $(\Omega, \mathscr{F}, P)$，假定 $\mathscr{F}_1, \mathscr{F}_2, \cdots, \mathscr{F}_n$ 是相互独立的子事件 σ 域。那么 $\mathscr{F}_1, \mathscr{F}_2, \cdots, \mathscr{F}_n$ 在子概率空间 $(\Omega, \mathscr{F}_1 \otimes \mathscr{F}_2 \otimes \cdots \otimes \mathscr{F}_n, P)$ 中也相互独立。并且 $P(\mathscr{F}_1 \otimes \mathscr{F}_2 \otimes \cdots \otimes \mathscr{F}_n)$ 由 n 个概率测度 $P(\mathscr{F}_i)(i=1,2,\cdots,n)$ 惟一地确定。

与两个子试验独立的情况类似，推论 2 和 3 反映 n 个独立的子试验是用何种方式存在于 $(\Omega, \mathscr{F}, P)$ 中。粗略地说，如果 $(\Omega, \mathscr{F}, P)$ 含有 n 个相互独立的事件，那么它包含 n 个相互独立的子试

验，从而包含一个与 n 维独立乘积概率空间本质上类似的子概率空间 $(\Omega,\mathscr{F}_1\otimes\mathscr{F}_2\otimes\cdots\otimes\mathscr{F}_n,P)$。

定理 2.9.5 设 $(\Omega,\mathscr{F},P)$ 是概率空间，$A_i\in\mathscr{F}(i=1,2,\cdots,n)$。那么 n 个事件 $A_1,A_2,\cdots,A_n$ 相互独立的充分必要条件是对任意的 $k(1\leqslant k\leqslant n)$ 成立

$$P(A_{n_1}A_{n_2}\cdots A_{n_k})=P(A_{n_1})P(A_{n_2})\cdots P(A_{n_k}) \quad (2.9.36)$$

注 (2.9.36)共包含 2^n-n-1 个等式。它们是

$$\left.\begin{array}{ll}P(A_{n_1}A_{n_2})=P(A_{n_1})P(A_{n_2}) & (\text{共 } C_n^2 \text{ 个})\\ P(A_{n_1}A_{n_2}A_{n_3})=P(A_{n_1})P(A_{n_2})P(A_{n_3}) & (\text{共 } C_n^3 \text{ 个})\\ \cdots\cdots & \\ P(A_{n_1}A_{n_2}\cdots A_{n_n})=P(A_{n_1})P(A_{n_2})\cdots P(A_{n_n}) & (\text{共 } C_n^n \text{ 个})\end{array}\right\} \quad (2.9.37)$$

如果其中的第一组等式成立，那么 $A_1,A_2,\cdots,A_n$ 中任意的两个事件独立。众所周知，两两独立不能推出 $A_1,A_2,\cdots,A_n$ 相互独立。因为在一般情形下用(2.9.37)的第一组等式不能推出(2.9.37)的其他等式。

证明 必要性：假定 $A_1,A_2,\cdots,A_n$ 相互独立。那么存在相互独立的子试验 $(\Omega,\mathscr{F}_1),(\Omega,\mathscr{F}_2),\cdots,(\Omega,\mathscr{F}_n)$ 使得 $A_i\in\mathscr{F}_i(i=1,2,\cdots,n)$。于是，对 $A_{n_1},A_{n_2},\cdots,A_{n_k},A_{n_j}=\Omega(k+1\leqslant j\leqslant n)$ 应用定理 2.9.4 即得(2.9.36)。

充分性：假定(2.9.36)成立。首先用它推出 2^n 个等式

$$\left.\begin{array}{ll}P(A_{n_1}A_{n_2}\cdots A_{n_n})=P(A_{n_1})P(A_{n_2})\cdots P(A_{n_n}) & (\text{共 } C_n^0 \text{ 个})\\ \cdots\cdots & \\ P(A_{n_1}\cdots A_{n_k}\overline{A}_{n_{k+1}}\cdots\overline{A}_{n_n})=P(A_{n_1})\cdots P(A_{n_k})P(\overline{A}_{n_{k+1}})\cdots P(\overline{A}_{n_n}) & \\ & (\text{共 } C_n^k \text{ 个})\\ \cdots\cdots & \\ P(\overline{A}_{n_1}\overline{A}_{n_2}\cdots\overline{A}_{n_n})=P(\overline{A}_{n_1})P(\overline{A}_{n_2})\cdots P(\overline{A}_{n_n}) & (\text{共 } C_n^n \text{ 个})\end{array}\right\} \quad (2.9.38)$$

事实上,(2.9.38)的第一组等式是(2.9.37)的最后一组等式。记 $B=A_{n_1}A_{n_2}\cdots A_{n_{n-1}}$,应用(2.9.37)得

$$P(BA_{n_n}) = P(A_{n_1})\cdots P(A_{n_{n-1}})P(A_{n_n}) = P(B)P(A_{n_n})$$

由(2.9.30)得出

$$P(B\overline{A}_{n_n}) = P(B)P(\overline{A}_{n_n}) = P(A_{n_1})\cdots P(A_{n_{n-1}})P(\overline{A}_{n_n})$$

得证(2.9.38)的第二组等式。用同样的方法可证第三组等式。依此类推得出(2.9.38)。

现在令

$$\mathscr{F}_i = \{\varnothing, A_i, \overline{A}_i, \Omega\} \quad (i = 1,2,\cdots,n) \qquad (2.9.39)$$

(2.9.38)保证 $(\Omega,\mathscr{F}_1),(\Omega,\mathscr{F}_2),\cdots,(\Omega_n,\mathscr{F}_n)$ 是 n 个相互独立的子试验。故 $A_1,A_2,\cdots,A_n$ 相互独立。 **证毕**

推论 假定 n 个事件 $A_1,A_2,\cdots,A_n$ 相互独立,那么其中任意的 m 个事件也相互独立($2\leqslant m\leqslant n-1$)。

2.9.5 任意个子试验的相互独立性

用 T 表示无限指标集。

定义 2.9.6 给定概率空间 $(\Omega,\mathscr{F},P)$,设 $(\Omega,\mathscr{F}_t)$ 是子随机试验($t\in T$)。如果子随机试验族 $\{(\Omega,\mathscr{F}_t)\mid t\in T\}$ 中任意有限个子试验皆相互独立,则称子试验族 $\{(\Omega,\mathscr{F}_t)\mid t\in T\}$ 相互独立,简称为子事件 σ 域族 $\{\mathscr{F}_t\mid t\in T\}$ 相互独立。

定义 2.9.7 给定概率空间 $(\Omega,\mathscr{F},P)$,设 $A_t\in\mathscr{F}(t\in T)$。如果存在相互独立的子随机试验族 $\{(\Omega,\mathscr{F}_t)\mid t\in T\}$ 使得 $A_t\in\mathscr{F}_t(t\in T)$,则称事件族 $\{A_t\mid t\in T\}$ 相互独立。

下面的两个定理是两个定义的直接推论。

定理 2.9.6 设 $(\Omega,\mathscr{F},P)$ 是概率空间,$\mathscr{F}_t$ 是 $\mathscr{F}$ 的子事件 σ 域($t\in T$)。那么下列 3 个结论相互等价:

(1) 子事件 σ 域族 $\{\mathscr{F}_t\mid t\in T\}$ 相互独立。

(2) $\{\mathscr{F}_t|t\in T\}$的任意子族$\{\mathscr{F}_t|t\in T'\}$相互独立，其中$T'$是$T$的任意子集。

(3) 对任意有限个$t_1,t_2,\cdots,t_n\in T$和$A_{t_1}\in\mathscr{F}_{t_1},A_{t_2}\in\mathscr{F}_{t_2},\cdots,A_{t_n}\in\mathscr{F}_{t_n}$成立

$$P(A_{t_1}A_{t_2}\cdots A_{t_n})=P(A_{t_1})P(A_{t_2})\cdots P(A_{t_n})\qquad(2.9.40)$$

推论 1 独立乘积概率空间$\left(\prod_{t\in T}\Omega_t,\prod_{t\in T}\mathscr{F}_t,\prod_{t\in T}P_t\right)$中子试验族$\left\{\left(\prod_{t\in T}\Omega_t,\mathscr{F}_s\right)\Big|s\in T\right\}$相互独立。

推论 2 设$(\Omega,\mathscr{F},P)$是概率空间，$\{\mathscr{F}_t|t\in T\}$是相互独立的子事件σ域族。那么在子概率空间$(\Omega,\bigotimes_{t\in T}\mathscr{F}_t,P)$上$\{\mathscr{F}_t|t\in T\}$也是相互独立的子事件$\sigma$域族，并且$P(\bigotimes_{t\in T}\mathscr{F}_t)$由概率测度族$\{P(\mathscr{F}_t)|t\in T\}$惟一确定。

与n个子试验相互独立的情形类似，推论1和2反映相互独立的子试验用何种方式存在于$(\Omega,\mathscr{F},P)$中。粗略地说，如果$(\Omega,\mathscr{F},P)$含有相互独立的事件族$\{A_t|t\in T\}$，那么它含有以T为指标的相互独立的子试验族$\{(\Omega,\mathscr{F}_t)|t\in T\}$，从而包含一个与$T$维独立乘积概率空间本质上类似的子概率空间$(\Omega,\bigotimes_{t\in T}\mathscr{F}_t,P)$。

定理 2.9.7 设$(\Omega,\mathscr{F},P)$是概率空间，$A_t\in\mathscr{F}(t\in T)$。那么下列4个结论相互等价：

(1) 事件族$\{A_t|t\in T\}$相互独立。

(2) $\{A_t|t\in T\}$的任意子族$\{A_t|t\in T'\}$相互独立，其中T'是T的子集。

(3) $\{A_t|t\in T\}$中任意有限个事件相互独立。

(4) 对任意有限个$t_1,t_2,\cdots,t_n\in T$成立

$$P(A_{t_1}A_{t_2}\cdots A_{t_n})=P(A_{t_1})P(A_{t_2})\cdots P(A_{t_n})\qquad(2.9.41)$$

2.9.6 条件概率值的基本定理

定理 2.9.8(乘法定理) 给定概率空间$(\Omega,\mathscr{F},P)$和$A_t\in\mathscr{F}$

$(i=1,2,\cdots,n)$。如果 $P(A_1A_2\cdots A_n)>0$,则

$$P(A_1A_2\cdots A_n)$$
$$=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{2.9.42}$$

证明 由于 $P(A_1)\geqslant P(A_1A_2)\geqslant\cdots\geqslant P(A_1A_2\cdots A_n)>0$,故(2.9.42)右方的条件概率皆有意义。现在应用(2.9.16)把条件概率值代入即知(2.9.42)成立。 **证毕**

定理 2.9.9(全概率定理) 给定概率空间$(\Omega,\mathscr{F},P)$和 Ω 的划分 $\mathscr{H}=\{H_i|i\in I\}$。假定 $H_i\in\mathscr{F},P(H_i)>0(i\in I)$。那么对任意的 $A\in\mathscr{F}$ 成立

$$P(A)=\sum_{i\in I}P(A|H_i)P(H_i) \tag{2.9.43}$$

证明 由于诸 $AH_i(i\in I)$ 是有限或可列个互不相容的事件,并且 $\bigcup_{i\in I}AH_i=A$,故

$$P(A)=\sum_{i\in I}P(AH_i)$$

对 $P(AH_i)$应用乘法定理即得(2.9.43)。 **证毕**

定理 2.9.10(贝叶斯定理) 在定理 2.9.9 的条件下,对任意的 $A\in\mathscr{F}^*$(即 $A\in\mathscr{F},P(A)>0$)成立

$$P(H_k|A)=\frac{P(A|H_k)P(H_k)}{\sum_{i\in I}P(A|H_i)P(H_i)}\quad(k\in I) \tag{2.9.44}$$

证明 我们有 $P(H_k|A)=\dfrac{P(AH_k)}{P(A)}(k\in I)$。对分子应用乘法定理,对分母应用全概率定理即得(2.9.44)。 **证毕**

习惯上,分别称(2.9.42),(2.9.43)和(2.9.44)为乘法公式,全概率公式和贝叶斯公式。

2.9.7 注记

(1) 独立性与现实世界的联系

现实世界中存在许许多多毫无联系的实验 $\mathscr{E}_t(t\in T)$。每个 $\mathscr{E}_t$ 产生一组因果关系：

$$\text{实验 } \mathscr{E}_t(\text{在划定条件 } \mathscr{C}_t \text{ 下}) \Longrightarrow (\Omega_t, \mathscr{F}_t, P_t) \tag{2.9.45}$$

如果把一些这样的实验人为地凑在一起，形成一个新实验 $\mathscr{E}'$。假定 $\underset{t\in T}{\circ}\Omega_t$ 是真交叉划分（在应用中通过 Ω_t 的适当选取总能达到目的），那么 $\mathscr{E}'$ 产生因果关系[1]

$$\text{实验 } \mathscr{E}'(\text{在独立条件族 } \mathscr{C}_t, t\in T \text{ 下}) \Longrightarrow \left(\prod_{t\in T}\Omega_t, \prod_{t\in T}\mathscr{F}_t, \prod_{t\in T}P_t\right) \tag{2.9.46}$$

假定 $\underset{t\in T}{\circ}\Omega_t$ 仅是交叉划分，这时只能得到

$$\text{实验 } \mathscr{E}'(\text{条件同上}) \Longrightarrow \left(\underset{t\in T}{\circ}\Omega_t, \bigotimes_{t\in T}\mathscr{F}_t, P\right) \tag{2.9.47}$$

类似于定理 2.9.6 后的议论，直观上可以认为 $\left(\underset{t\in T}{\circ}\Omega_t, \bigotimes_{t\in T}\mathscr{F}_t, P\right)$ 本质上类似于一个独立乘积概率空间。

现实中人们经常研究一个大实验 $\mathscr{E}$。它产生因果关系

$$\text{实验 } \mathscr{E}(\text{在划定条件 } \mathscr{C} \text{ 下}) \Longrightarrow (\Omega, \mathscr{F}, P) \tag{2.9.48}$$

之所以称 $\mathscr{E}$ 为大实验，是指它包含许多子实验，其中有相互毫无关联的实验 $\mathscr{E}_t(t\in T)$。这个"因"反映在概率空间中的"果"是结论：$(\Omega, \mathscr{F}, P)$ 含有子概率空间 $\left(\Omega, \bigotimes_{t\in T}\mathscr{F}_t, P\right)$。

定理 2.9.1，定理 2.9.4，定理 2.9.6 的推论表明，只要 $(\Omega, \mathscr{F}, P)$ 中存在相互独立的事件族 $\{A_t | t\in T\}$，那么 $(\Omega, \mathscr{F}, P)$ 便来自(2.9.48)中的大实验且 $A_t \in \mathscr{F}_t$。这就是事件相互独立的本质。

(2) 全概率公式如何推广？

在全概率公式(2.9.43)中 A 可以是 $\mathscr{F}$ 中任意的事件，$P(A|H_i)$ 是条件概率 $P(A|\sigma(\mathscr{H}))$ 的一些数值。因此全概率定理可以

---

① 这是独立实验原理，它将被抽象为公理 VI_5。

叙述为：概率测度 $P(\mathscr{F})$ 由概率测度 $P(\sigma(\mathscr{H}))$ 和条件概率 $P(\mathscr{F}|\sigma(\mathscr{H}))$ 用公式(2.9.43)惟一地确定。

为了显示全概率公式的巨大潜力，必须把定理 2.9.9 推广到对任意的子试验 $(\Omega,\mathscr{F}_2)$ 成立。即得到结论：$P(\mathscr{F})$ 由 $P(\mathscr{F}_2)$ 和条件概率 $P(\mathscr{F}|\mathscr{F}_2)$ 惟一地确定[1]，并且找到(2.9.43)的一般形式。

下面的讨论富有启发性，但是非常不严格。让我们用 Ω 的最细划分 $\Omega=\{\langle\omega\rangle|\omega\in\Omega\}$ 代替划分 $\mathscr{H}$，并且形式地写出(2.9.43)的推广式

$$P(A)\sim\sum_{\omega\in\Omega}P(A|\omega)P(\langle\omega\rangle) \qquad (*)$$

这种形式工作带来两个严重的问题：

① 样本点可以不是事件，即使是事件也可以不属于 $\mathscr{F}$，此时 $P(\langle\omega\rangle)$ 无意义。退一步说，即使 $\langle\omega\rangle\in\mathscr{F}$，使 $P(\langle\omega\rangle)=0$ 的样本点集合可以是正概率的，而且往往概率为 1。例如，一般的柯氏型概率空间就是这样的。

② 对①中所提及的样本点 ω，条件概率 $P(A|\omega)$ 无意义。

解决这两个问题的可能途径是：用子试验 $(\Omega,\mathscr{F}_2)$ 代替最细划分 Ω；引进 $\mathscr{F}_2$ 中的"无穷小事件"$\mathrm{d}\omega$，并且用 $P(\mathrm{d}\omega)$ 代替 $P(\langle\omega\rangle)$；$P(A|\omega)$ 则认为是 $(\Omega,\mathscr{F}_2)$ 上的点函数 $g(\omega)$，但要求它能惟一地确定条件概率 $P(A|\mathscr{F}_2)$。如果能实现这三件事情，依据微积分的思想方法我们有希望得到

$$P(A)\sim\int_\Omega g(\omega)P(\mathrm{d}\omega) \qquad (**)$$

$(**)$式的确是一个可证明的数学等式。事实上，它是测度论中的拉东—尼科迪姆(Radon-Nikodym)定理。

于是，我们从"非常不严格的讨论"中得到 3 点收获：① $(**)$式原来是一个已被证明的式子；② 条件概率 $P(A|\mathscr{F}_2)$ 可以用点函数 $g(\omega)$ 确定；③ $(**)$式可作为确定 $g(\omega)$ 的工具。这三点正

① 这个结论很容易地由(2.9.16)和(2.9.21)推出。

是产生出 2.10 节内容的源泉。

(3) 为了显示下节定理的概率内容,在此处介绍:

拉东--尼科迪姆定理:给定概率空间$(\Omega,\mathscr{F}_2,P)$。假定$\nu(A)$,$A\in\mathscr{F}_2$是满足公理IV_3的集函数,并且关于测度$P(\mathscr{F}_2)$绝对连续(即$P(A)=0$蕴含$\nu(A)=0$)。那么在$(\Omega,\mathscr{F}_2)$上存在$\mathscr{F}_2$可测的点函数$g(\omega)$,$\omega\in\Omega$,使得对任意的$A\in\mathscr{F}_2$成立

$$\nu(A)=\int_A g(\omega)P(\mathrm{d}\omega) \qquad (2.9.49)$$

并且在不计$\mathscr{F}_2$中零概率集意义下$g(\omega)$是惟一的。

证明见文献[16]。为了不出现陌生的术语和记号,此处把测度空间限制为概率空间$(\Omega,\mathscr{F}_2,P)$。

2.10 随机试验上的点函数(续)

和概率测度一样,人们希望各种形式的条件概率能够用点函数确定。寻找这样的点函数再次显示限制是一种有内涵的数学运算。

2.10.1 条件密度 $p(A|\mathscr{F}_2)(\omega)$

事件A关于子试验$(\Omega,\mathscr{F}_2)$的条件概率$P(A|\mathscr{F}_2)$是式(2.9.22)规定的一组条件概率值。从集函数的角度很难看出这些值之间存在联系。但是,这些值之间的确存在紧密的联系,并且可以用点函数表达出这种联系(参看上节节末注记2))。

定义 2.10.1 给定概率空间$(\Omega,\mathscr{F},P)$。假定$A\in\mathscr{F}$,$(\Omega,\mathscr{F}_2)$是子试验。称定义在$(\Omega,\mathscr{F})$上的点函数$g(\omega)$,$\omega\in\Omega$是条件概率$P(A|\mathscr{F}_2)$的条件密度,如果g具有性质:

(1) $g(\omega)$是$\mathscr{F}_2$可测函数;

(2) $g(\omega)$用公式

$$P(A|B)=\frac{1}{P(B)}\int_B g(\omega)P(\mathrm{d}\omega) \quad (B\in\mathscr{F}_2^*) \qquad (2.10.1)$$

确定 $P(A|\mathscr{F}_2)$。

今后，用专有符号 $p(A|\mathscr{F}_2)(\omega)$ 表示 $P(A|\mathscr{F}_2)$ 的条件密度，也可称之为 A 关于 $\mathscr{F}_2$ 的条件密度。

定理 2.10.1 设 $(\Omega,\mathscr{F},P)$ 是概率空间。那么其上任意的条件概率 $P(A|\mathscr{F}_2)$ 皆存在条件密度，并且在不计 $\mathscr{F}_2$ 中零概率集的意义下是惟一的。

证明 定理的条件含有 $(\Omega,\mathscr{F}_2,P)$ 是概率空间。对任意固定的 $A\in\mathscr{F}$，容易验证集函数 $P(AB)$，$B\in\mathscr{F}_2$ 满足公理 Ⅳ_3 和关于 $P(\mathscr{F}_2)$ 绝对连续。由拉东—尼科迪姆定理得出，在不计 $\mathscr{F}_2$ 中零概率集的意义下存在惟一的 $\mathscr{F}_2$ 可测函数 $g(\omega)$，$\omega\in\Omega$ 使得

$$P(AB)=\int_B g(\omega)P(\mathrm{d}\omega)\quad (B\in\mathscr{F}_2)\qquad (2.10.2)$$

当 $B\in\mathscr{F}_2^*$ 时两边除以 $P(B)$ 即得(2.10.1)。得证 g 是 $P(A|\mathscr{F}_2)$ 的条件密度，并且是惟一的。 **证毕**

关于新概念作 5 点说明：

(1) 在理论推导和概率计算中用 $p(A|\mathscr{F}_2)$ 代替 $P(A|\mathscr{F}_2)$ 时带来巨大的方便，因此条件密度在现代概率论中占据重要的地位。

在现今的概率论中由于没有引进我们的条件概率 $P(A|\mathscr{F}_2)$，因此人们把 $p(A|\mathscr{F}_2)$ 称为"A 关于 $\mathscr{F}_2$ 的条件概率"且记为"$P(A|\mathscr{F}_2)$"。为了贴近概念的本质，我们使用了新名称和新符号，希望公众能采纳。

(2) 条件密度与条件概率不同，它没有一般性的直观解释。现今流行的，但不成文的直观解释为："$p(A|\mathscr{F}_2)(\omega)$ 是子试验 $(\Omega,\mathscr{F}_2)$ 进行时样本点 ω 出现的条件下事件 A 的条件概率值。"[①] 事实上，"$(\Omega,\mathscr{F}_2)$ 进行时样本点 ω 出现"是条件 $\langle\omega\rangle\in\mathscr{F}_2$。此时若 $P(\langle\omega\rangle)>0$，则在(2.10.1)中取 $B=\langle\omega\rangle$ 得到

---

① 它或许是直观解释的首次书面表达。原因可能是原先的概率论中没有"子试验 $(\Omega,\mathscr{F}_2)$ 进行时样本点 ω 出现"的表述方式。因为如果改为"样本点 ω 出现的条件下事件 A 的条件概率"，那么直观解释就不正确了，见正文的议论。

$$p(A|\mathscr{F}_2)(\omega)=P(A|\langle\omega\rangle) \tag{2.10.3}$$

因此直观解释正确；若 $P(\langle\omega\rangle)=0$，(2.10.3)右方无意义，左方的值无关紧要(因为 $p(A|\mathscr{F}_2)(\omega)$可以在 $\mathscr{F}_2$ 的一个零概率集上无意义)，因此袭用解释也无不可。

务须注意，当 $\langle\omega\rangle\overline{\in}\mathscr{F}_2$ 时直观解释失效。事实上，当 $\langle\omega\rangle\overline{\in}\mathscr{F}_2$，$\langle\omega\rangle\in\mathscr{F}$ 且 $P(\langle\omega\rangle)>0$ 时经常成立(见本节的例 1)

$$p(A|\mathscr{F}_2)(\omega)\neq P(A|\langle\omega\rangle) \tag{2.10.4}$$

另外，直观解释中"子试验$(\Omega,\mathscr{F}_2)$进行"必须强调。因为对不同的子试验$(\Omega,\mathscr{F}_3)$，使不等式

$$p(A|\mathscr{F}_2)(\omega)\neq p(A|\mathscr{F}_3)(\omega) \tag{2.10.5}$$

成立的样本点 ω 不仅可以是正概率集，而且往往是概率为 1 的集合(见本节的例 1)。

(3) 设$(\Omega,\mathscr{F}_2)$和$(\Omega,\mathscr{F}_3)$是两个子试验，并且 $\mathscr{F}_3\subset\mathscr{F}_2$。作为集函数的条件概率成立

$$P(A|\mathscr{F}_3)=P(A|\mathscr{F}_2)\upharpoonright\mathscr{F}_3 \tag{2.10.6}$$

但是，作为点函数的条件密度 $p(A|\mathscr{F}_2)$和 $p(A|\mathscr{F}_3)$却是完全不同的可测函数(参看(2.10.5)和例 1)。这个事实说明限制是一种有内涵的数学运算。

(4) 当我们取不同的事件 A 时得到关于子试验$(\Omega,\mathscr{F}_2)$的条件概率族

$$\{P(A|\mathscr{F}_2)|A\in\mathscr{F}\}^{①} \tag{2.10.7}$$

和条件密度族

$$\{p(A|\mathscr{F}_2)(\omega)|A\in\mathscr{F}\} \tag{2.10.8}$$

对族(2.10.8)中的点函数，可以讨论它们的取值，正负性和四则运算。应用积分知识容易由定义 2.10.1 推出

---

① 此即试验$(\Omega,\mathscr{F})$关于子试验$(\Omega,\mathscr{F}_2)$的条件概率 $P(\mathscr{F}|\mathscr{F}_2)$。

① 对任意的 $A\in\mathscr{F}$ 有 $p(A|\mathscr{F}_2)(\omega)\geqslant 0\quad (P(\mathscr{F}_2)\text{-a.s.})$

$$(2.10.9)$$

② $p(\Omega|\mathscr{F}_2)(\omega)=1\quad (P(\mathscr{F}_2)\text{-a.s.})$ (2.10.10)

③ 对 $\mathscr{F}$ 中互不相容的事件 $A_i(i\in I)$ 成立

$$p\Big(\bigcup_{i\in I}A_i\,|\,\mathscr{F}_2\Big)(\omega)=\sum_{i\in I}p(A_i|\mathscr{F}_2)(\omega)\quad (P(\mathscr{F}_2)\text{-a.s.})\tag{2.10.11}$$

其中"$P(\mathscr{F}_2)$-a.s."表示前方的等式或不等式可以在 $\mathscr{F}_2$ 中的一个零概率集上不成立，甚至无意义。为了书写方便，人们通常把 $P(\mathscr{F}_2)$-a.s. 省略。

应当指出，(2.10.9)～(2.10.11)与公理Ⅳ$_1$～Ⅳ$_3$ 中的相应式在外表上非常类似，但却有本质的不同。这里的 $p(A|\mathscr{F}_2)$，$p(\Omega|\mathscr{F}_2)$，$p(A_i|\mathscr{F}_2)$等仅仅是函数对应律的符号。不同的 A，Ω，A_i 表明它们是不同的函数，不能把它们误认为自变量。

(5) 当各种各样的子试验都可以进行时我们得到更大的条件概率族

$$\{P(A|\mathscr{F}_2)\,|\,A\in\mathscr{F},\mathscr{F}_2\subset\mathscr{F},\mathscr{F}_2\text{ 是事件 }\sigma\text{ 域}\}\tag{2.10.12}$$

和更大的条件密度族

$$\{p(A|\mathscr{F}_2)(\omega)\,|\,A\in\mathscr{F},\mathscr{F}_2\subset\mathscr{F},\mathscr{F}_2\text{ 是事件 }\sigma\text{ 域}\}\tag{2.10.13}$$

族(2.10.13)中点函数除具有性质(2.10.9)～(2.10.11)外还有许多重要的性质，它们在现代概率论中占据非常重要的地位。由于这些函数是第 3 章中的随机变量，它们的深入研究将在概率论公理化后进行。

2.10.2 条件密度—概率

定理 2.10.1 的结论又可写成：

定理 2.10.2 给定概率空间$(\Omega,\mathscr{F},P)$。假定$(\Omega,\mathscr{F}_1)$和

$(\Omega,\mathscr{F}_2)$是两个子试验，那么条件概率（测度族）$P(\mathscr{F}_1|\mathscr{F}_2)$和条件密度族

$$\{p(A|\mathscr{F}_2)(\omega)\,|\,A\in\mathscr{F}_1\} \tag{2.10.14}$$

相互惟一确定，并且满足

$$P(A|B)=\frac{1}{P(B)}\int_B p(A|\mathscr{F}_2)(\omega)P(\mathrm{d}\omega)$$

$$(A\in\mathscr{F}_1,B\in\mathscr{F}_2^*) \tag{2.10.15}$$

再次提醒注意，$p(A|\mathscr{F}_2)(\omega)$中的$p(A|\mathscr{F}_2)$仅仅是对应律的符号。不能把函数族(2.10.14)误认为二元函数$p(A|\mathscr{F}_2)(\omega)$，$(A,\omega)\in\mathscr{F}_1\times\Omega$。事实上，对固定的$A$，由于$p(A|\mathscr{F}_2)(\omega)$是$\mathscr{F}_2$可测函数，因此它的准确的解析表达式为

$$p(A|\mathscr{F}_2)(\omega),\quad \omega\in\Omega\backslash N_A \tag{2.10.16}$$

其中$N_A\in\mathscr{F}_2$且$P(N_A)=0$。由此得出(2.10.14)只能产生出一个二元集—点函数

$$g(A,\omega)=p(A|\mathscr{F}_2)(\omega),A\in\mathscr{F}_1,\omega\in\Omega\Big\backslash\bigcup_{A\in\mathscr{F}_1}N_A \tag{2.10.17}$$

一般说来，$\mathscr{F}_1$含有不可列个事件。作为集合的不可列并，$\bigcup\limits_{A\in\mathscr{F}_1}N_A$可以不是事件；即使是事件，也不必是零概率的。

但是，我们的确希望(2.10.14)能够成为一个定义在$\mathscr{F}_1\times\Omega$上的集—点函数。若如此，我们便可以用这个集—点函数确定条件概率$P(\mathscr{F}_1|\mathscr{F}_2)$，为研究$P(\mathscr{F}_1|\mathscr{F}_2)$带来许多的方便。

定义 2.10.2 称二元集—点函数$g(A,\omega)$，$(A,\omega)\in\mathscr{F}_1\times\Omega$为条件概率$P(\mathscr{F}_1|\mathscr{F}_2)$的条件密度—概率，如果它具有下列3个性质：

(1) 对固定的A，g关于ω是$\mathscr{F}_2$可测函数；

(2) 对固定的ω，g关于A是$(\Omega,\mathscr{F}_1)$上的概率测度；

(3) $g(A,\omega)$用公式

$$P(A|B) = \frac{1}{P(B)}\int_B g(A,\omega)P(\mathrm{d}\omega) \quad (A \in \mathscr{F}_1, B \in \mathscr{F}_2)$$

(2.10.18)

确定 $P(\mathscr{F}_1|\mathscr{F}_2)$。

今后，用专有符号 $p(A|\omega;\mathscr{F}_2)$ 表示 $g(A,\omega)$。如果要强调定义域，则可使用符号 $p(\mathscr{F}_1|\Omega;\mathscr{F}_2)$，并简称 $p(A|\omega;\mathscr{F}_2)$ 为 $\mathscr{F}_1$ 关于 $\mathscr{F}_2$ 的条件密度—概率[①]。与条件密度不同，条件密度—概率可能存在，也可能不存在。

定理 2.10.3 给定概率空间 $(\Omega,\mathscr{F},P)$。如果(2.10.16)中的 N_A 满足条件

$$\bigcup_{A\in\mathscr{F}_1} N_A \in \mathscr{F}_2, P\Big(\bigcup_{A\in\mathscr{F}_1} N_A\Big) = 0 \qquad (2.10.19)$$

那么条件概率 $P(\mathscr{F}_1|\mathscr{F}_2)$ 存在条件密度—概率。

证明 应用(2.10.19)可以把可测版本(2.10.16)改为

$$p(A|\mathscr{F}_2)(\omega), \omega \in \Omega\backslash N \qquad (2.10.20)$$

其中 $N=\bigcup_{A\in\mathscr{F}_1} N_A$。于是可以构造二元集—点函数

$$g(A,\omega)=\begin{cases} p(A|\mathscr{F}_2)(\omega) & (A,\omega)\in\mathscr{F}_1\times\Omega\backslash N \\ \text{任意赋值} & (A,\omega)\in\mathscr{F}_1\times N \end{cases} \qquad (2.10.21)$$

显然 $g(A,\omega)$ 具有定义 2.10.2 中性质(1)。性质(3)由定理 2.10.2得出。当 $\omega\in\Omega\backslash N$ 时式(2.10.9)～(2.10.11)成为性质(2)；当 $\omega\in N$ 时可以要求“任意赋值”满足(2)。得证 $g(A,\omega)$ 是条件密度—概率 $p(A|\omega;\mathscr{F}_2)$。 **证毕**

容易看出，当 $\mathscr{F}_1$ 是有限集时条件(2.10.19)成立。可以证明，如果概率空间 $(\Omega,\mathscr{F},P)$ 中 Ω 是完备可分距离空间，$\mathscr{F}$ 是其上的波莱尔域，那么 $(\Omega,\mathscr{F},P)$ 上任意的条件概率 $P(\mathscr{F}_1|\mathscr{F}_2)$ 皆存在

---

① 在现代概率论中称 $p(A|\omega;\mathscr{F}_2)$ 为正则条件概率。考虑到贴近概念本质的名称使思维流畅，我们使用了条件密度—概率这个名称。

条件密度—概率。这个结论的一种特殊情形是

定理 2.10.4 设$(\Omega,\mathscr{F},P)$是离散型，n 维或可列维柯氏概率空间。那么其上的条件概率 $P(\mathscr{F}_1|\mathscr{F}_2)$ 皆存在条件密度—概率。

证明见文献[22]。

至此，我们解决了把(2.9.21)和(2.9.22)中变元 B 转化为变元 ω 的问题。本节后半部分讨论把(2.9.21)和(2.9.23)中变元 A 转化为变元 ω 的问题。和概率测度一样，这个问题只能在几类典型的概率空间中得到解答。

2.10.3 离散型概率空间上的点函数

在本段中假定$(\Omega,\mathscr{F},P)$是离散型概率空间。其中

$$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n,\cdots\}=\{\omega_n|n\in I\} \qquad (2.10.22)$$

$P(\mathscr{F})$由分布律 $f(\omega)$，$\omega\in\Omega$ 产生，这里

$$f:\begin{pmatrix}\omega_1 & \omega_2 & \cdots & \omega_n & \cdots\\ f_1 & f_2 & \cdots & f_n & \cdots\end{pmatrix} \qquad (2.10.23)$$

并且约定，本段的讨论以分布律，而不以概率测度作为研究的出发点。即是说，即使 $P(\mathscr{F})$存在其他的分布律(参看定理2.8.1)，我们也不讨论它。虽然其他分布律的存在不改变解答的本质，但会使叙述冗长。

(1) 条件分布律 $f(\omega|B)$

定义 2.10.3 给定离散型概率空间$(\Omega,\mathscr{F},P)$和 $P(\mathscr{F})$的分布律 (2.10.23)。又设 $B\in\mathscr{F}$，$P(B)>0$。则称$(\Omega,\mathscr{F})$上的点函数

$$f(\omega_n|B)=\frac{f_n\chi_B(\omega_n)}{P(B)},\ \omega_n\in\Omega \qquad (2.10.24)$$

即

$$f(\omega|B):\begin{pmatrix}\omega_1 & \omega_2 & \cdots & \omega_n & \cdots\\ \dfrac{f_1\chi_B(\omega_1)}{P(B)} & \dfrac{f_2\chi_B(\omega_2)}{P(B)} & \cdots & \dfrac{f_n\chi_B(\omega_n)}{P(B)} & \cdots\end{pmatrix} \qquad (2.10.25)$$

为分布律 f 关于事件 B 的条件分布律，简称为关于 B 的条件分布律。其中 $\chi_B(\omega)$ 是 B 的示性函数且 $P(B)=\sum\limits_{n:\omega_n\in B} f_n$。

定理 2.10.5 给定离散型概率空间 $(\Omega,\mathscr{F},P)$ 和 $P(\mathscr{F})$ 的分布律(2.10.23)。那么对任意的子试验 $(\Omega,\mathscr{F}_1)$，条件概率(测度) $P(\mathscr{F}_1|B)$ 由条件分布律 $f(\omega|B)$ 用公式

$$P(A|B)=\sum_{\omega\in A} f(\omega|B) \quad (A\in\mathscr{F}_1) \qquad (2.10.26)$$

确定。

证明 把(2.10.24)代入(2.10.26)的右方得到

$$\begin{aligned}\sum_{\omega\in A} f(\omega|B) &= \frac{1}{P(B)}\sum_{n:\omega_n\in A} f_n\chi_B(\omega_n) \\ &= \frac{1}{P(B)}\sum_{n:\omega_n\in AB} f_n = \frac{P(AB)}{P(B)}\end{aligned}$$

得证定理的结论。 **证毕**

(2) 条件密度 $p(A|\mathscr{F}_2)(\omega)$

设 $(\Omega,\mathscr{F}_2)$ 是 $(\Omega,\mathscr{F})$ 的子试验。由定理 2.5.3 和 2.5.6 得出，存在 $\mathscr{F}_2$ 中原子事件 $H_k'(k\geqslant 1)$ 使得

$$\mathscr{H}=\{H_k'|k\geqslant 1\} \qquad (2.10.27)$$

是 Ω 的划分且 $\mathscr{F}_2=\sigma(\mathscr{H})$。这时 $(\mathscr{H},\mathscr{F}_2)$ 是标准试验。

现在把使 $P(H_k')>0$ 的 H_k' 记为 $H_i(i\geqslant 1)$，把使 $P(H_k')=0$ 的 H_k' 记为 $N_j(j\geqslant 1)$。于是

$$\mathscr{H}=\{H_i,N_j|i\geqslant 1,j\geqslant 1\} \qquad (2.10.28)$$

又假定

$$H_i=\{\omega_{i1},\omega_{i2},\cdots\} \quad (i\geqslant 1) \qquad (2.10.29)$$

$$N=\bigcup_{j\geqslant 1} N_j=\{\omega_1',\omega_2',\cdots\} \qquad (2.10.30)$$

这时可以给出分布律(2.10.23)的另一种表达式

$$f:\begin{cases} H_1 & H_2 & \cdots & H_i & \cdots & N \\ \omega_{11}\omega_{12} \quad \cdots & \omega_{21}\omega_{22} \quad \cdots & \cdots & \omega_{i1}\omega_{i2} \quad \cdots & \cdots & \omega_1'\omega_2' \quad \cdots \\ f_{11}\, f_{12} \quad \cdots & f_{21}f_{22} \quad \cdots & \cdots & f_{i1}f_{i2} \quad \cdots & \cdots & 0 \quad 0 \quad \cdots \end{cases}$$

(2.10.31)

定理 2.10.6 给定离散型概率空间$(\Omega,\mathscr{F},P)$和子试验$(\Omega,\mathscr{F}_2)$。假定$\mathscr{F}_2=\sigma(\mathscr{H})$，那么对任意的$A\in\mathscr{F}$，条件概率$P(A|\mathscr{F}_2)$的条件密度为

$$p(A|\mathscr{F}_2)(\omega)=\sum_{i\geqslant 1}P(A|H_i)\chi_{H_i}(\omega) \quad (\omega\in\Omega) \quad (2.10.32)$$

即

$$P(A|\mathscr{F}_2):\begin{cases} H_1 & H_2 & \cdots & N \\ \omega_{11}\omega_{12}\cdots & \omega_{21}\omega_{22}\cdots & \cdots & \omega_1'\omega_2'\cdots \\ P(A|H_1) & P(A|H_2) & \cdots & 0\ 0\cdots \end{cases} \quad (2.10.33)$$

其中 $P(A|H_i)=\sum\limits_{j:\omega_{ij}\in A}f_{ij}\Big/\sum\limits_{j\geqslant 1}f_{ij} \quad (i=1,2,3,\cdots) \quad (2.10.34)$

证明 由于$\mathscr{F}_2=\sigma(\mathscr{H})$，故(2.10.32)定义的函数是$\mathscr{F}_2$可测函数。由积分知识得知，对具有分布律(2.10.23)的离散型概率空间，式(2.10.1)成为

$$P(A|B)=\frac{1}{P(B)}\sum_{n:\omega_n\in B}p(A|\mathscr{F}_2)(\omega_n)f_n \quad (B\in\mathscr{F}_2^*)$$

(2.10.35)

为完成证明，只须证(2.10.32)规定的点函数满足(2.10.35)。为此，把(2.10.32)代入(2.10.35)的和式中，我们有

$$\begin{aligned}\sum_{n:\omega_n\in B}\Big[\sum_{i\geqslant 1}P(A|H_i)\chi_{H_i}(\omega_n)\Big]f_n &= \sum_{i\geqslant 1}\Big[\sum_{n:\omega_n\in B}\chi_{H_i}(\omega_n)f_n\Big]P(A|H_i) \\ &= \sum_{i\geqslant 1}\Big[\sum_{n:\omega_n\in BH_i}f_n\Big]P(A|H_i) \\ &= \sum_{i\geqslant 1}P(BH_i)P(A|H_i) \\ &= \sum_{i\geqslant 1}P(AH_i)\frac{P(BH_i)}{P(H_i)}\end{aligned}$$

由 H_i 的原子性推出 $BH_i=H_i$ 或 $BH_i=\varnothing$。故

$$\text{上式} = \sum_{i:H_i\subset B} P(AH_i) = \sum_{i\geqslant 1} P(ABH_i) = P(AB)$$

得证(2.10.35)成立。 **证毕**

例 1 假定$(\Omega_5,\mathscr{F}_5,P)$是古典型概率空间,其中$(\Omega_5,\mathscr{F}_5)$是 2.4 节例 5 中随机试验。又设$(\Omega_5,\mathscr{F}_6)$和$(\Omega_5,\mathscr{F}_7)$分别是 2.4 节例 6 和例 7 中随机试验。这时,$(\Omega_5,\mathscr{F}_6)$和$(\Omega_5,\mathscr{F}_7)$是$(\Omega_5,\mathscr{F}_5)$的子试验。

令 $A=\{1,2,3\}(\in\mathscr{F}_5)$。应用定理 2.10.6 得出 A 关于 $\mathscr{F}_6$ 和 $\mathscr{F}_7$ 的条件密度分别为

$$p(A|\mathscr{F}_6):\begin{pmatrix} 1 & 3 & 5 & 2 & 4 & 6 \\ \frac{2}{3} & \frac{2}{3} & \frac{2}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \quad (2.10.36)$$

$$p(A|\mathscr{F}_7):\begin{pmatrix} 5 & 6 & 1 & 2 & 3 & 4 \\ 0 & 0 & \frac{3}{4} & \frac{3}{4} & \frac{3}{4} & \frac{3}{4} \end{pmatrix} \quad (2.10.37)$$

对退化试验$(\Omega,\mathscr{F}_0)$,A 关于 $\mathscr{F}_0$ 的条件密度为

$$p(A|\mathscr{F}_0):\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad (2.10.38)$$

本例可用于验证 2.10.1 段中的有关讨论。

(3) 条件密度—概率 $p(A|\omega;\mathscr{F}_2)$

定理 2.10.7 给定离散型概率空间$(\Omega,\mathscr{F},P)$和分布律(2.10.23)。假定$(\Omega,\mathscr{F}_1)$和$(\Omega,\mathscr{F}_2)$是子试验且 $\mathscr{F}_2=\sigma(\mathscr{H})$。那么条件概率 $P(\mathscr{F}_1|\mathscr{F}_2)$存在条件密度—概率。它是[①]

$$p(A|\omega;\mathscr{F}_2) = \sum_{i\geqslant 1} P(A|H_i)\chi_{H_i}(\omega)(A\in\mathscr{F}_1,\omega\in\Omega) \quad (2.10.39)$$

① 比较(2.10.39)与(2.10.32)得出对任意的 $A\in\mathscr{F}_1$ 成立

$$p(A|\mathscr{F}_2)(\omega)=p(A|\omega;\mathscr{F}_2),(\omega\in\Omega)$$

即

ω / A	H_1			H_2			N		
	ω_{11}	ω_{12}	$\cdots$	ω_{21}	ω_{22}	$\cdots$	ω_1'	ω_2'	$\cdots$
A_1	$P(A_1\|H_1)$	$P(A_1\|H_1)$	$\cdots$	$P(A_1\|H_2)$	$P(A_1\|H_2)$	$\cdots$	0	0	$\cdots$
A_2	$P(A_2\|H_1)$	$P(A_2\|H_1)$	$\cdots$	$P(A_2\|H_2)$	$P(A_2\|H_2)$	$\cdots$	0	0	$\cdots$
$\vdots$	……	……	$\cdots$	……	……		(也可不定义或任意赋值)		
一般	$P(\mathscr{F}_1\|H_1)$	$P(\mathscr{F}_1\|H_1)$	$\cdots$	$P(\mathscr{F}_1\|H_2)$	$P(\mathscr{F}_1\|H_2)$	$\cdots$			

(2.10.40)

其中$P(A|H_i)(A\in\mathscr{F}_1,i\geqslant 1)$由(2.10.34)规定。

证明 用$g(A,\omega)$表示(2.10.39)的右方。由于$\mathscr{F}_2=\sigma(\mathscr{H})$，故对固定的$A$,$g(A,\omega)$关于$\omega$是$\mathscr{F}_2$可测函数。当$\omega\overline{\in}N$时由定理2.10.6得出$g(A,\omega)$,$A\in\mathscr{F}_1$是$(\Omega,\mathscr{F}_1)$上的概率测度；当$\omega\in N$时我们或者不考虑它，或者通过“任意赋值”使$g(A,\omega)$,$A\in\mathscr{F}_1$是$(\Omega,\mathscr{F}_1)$上的概率测度。总之，$g(A,\omega)$满足定义2.10.2中的性质①和②。注意到对固定的$A\in\mathscr{F}_1$有$g(A,\omega)=p(A|\mathscr{F}_2)(\omega)$，应用定理2.10.6得$g(A,\omega)$也满足定义2.10.2中性质③。得证(2.10.39)是$P(\mathscr{F}_1|\mathscr{F}_2)$的条件密度—概率。 **证毕**

(4) 条件密度—分布律$f(\omega^*|\omega;\mathscr{F}_2)$

在定理2.10.7中取$\mathscr{F}_1=\mathscr{F}$，对固定的$\omega$得到$(\Omega,\mathscr{F})$上的概率测度$p(\mathscr{F}|\omega;\mathscr{F}_2)$①。用它的分布律可以引出条件密度—分布律的概念。由于应用(2.10.31)很容易写出这个分布律，于是采用下面的定义。

定义2.10.4 给定离散型概率空间$(\Omega,\mathscr{F},P)$。假定$(\Omega,\mathscr{F}_2)$是子试验且$\mathscr{F}_2=\sigma(\mathscr{H})$，$P(\mathscr{F})$有分布律(2.10.31)。则称$(\Omega,\mathscr{F})$上的二元点函数

---

① 按定义2.7.1下的脚注，它表示固定ω时的概率测度$p(A|\omega;\mathscr{F}_2)$，$A\in\mathscr{F}$。$p(\mathscr{F}_1|\Omega;\mathscr{F}_2)$是定理2.10.7中条件密度—概率的简写。

$$f(\omega^* \mid \omega;\mathscr{F}_2)=\begin{cases}\dfrac{f_{ij}\chi_{H_i}(\omega)}{\sum\limits_{k\geqslant 1} f_{ik}} & \omega^*=\omega_{ij},\omega\in\Omega(i\geqslant 1,j\geqslant 1)\\ 0 & \omega^*\in N,\omega\in\Omega\end{cases}$$

(2.10.41)

为分布律 f 关于 $\mathscr{F}_2$ 的条件密度一分布律，简称为关于 $\mathscr{F}_2$ 的条件密度一分布律。

注　条件密度一分布律 $f(\omega^* \mid \omega;\mathscr{F}_2)$ 的表格式为

ω \ ω^*	H_1: $\omega_{11}\omega_{12}$ …	H_2: $\omega_{21}\omega_{22}$ …	N: ω'_1 ω'_2 …	…
ω_{11}, ω_{12}, $\vdots$	$\dfrac{f_{1j}}{\sum\limits_{k\geqslant 1} f_{1k}}$	0		
ω_{21}, ω_{22}, $\vdots$	0	$\dfrac{f_{2j}}{\sum\limits_{k\geqslant 1} f_{2k}}$	0	0
$\vdots$		0		

(2.10.42)

定理 2.10.8　条件密度一分布律 $f(\omega^* \mid \omega;\mathscr{F}_2)$ 具有性质：

① 对固定的 ω^*，它关于 ω 是 $\mathscr{F}_2$ 可测函数；

② 对固定的 ω，它关于 ω^* 是 $(\Omega,\mathscr{F})$ 上的分布律；

③ 对任意的子试验 $(\Omega,\mathscr{F}_1)$，它用公式

$$p(A \mid \omega;\mathscr{F}_2)=\sum_{\omega^*\in A} f(\omega^* \mid \omega;\mathscr{F}_2)(A\in\mathscr{F}_1,\omega\in\Omega)$$

(2.10.43)

确定$\mathscr{F}_1$关于$\mathscr{F}_2$的条件密度—概率$p(\mathscr{F}_1|\Omega;\mathscr{F}_2)$。

证明 由于$\mathscr{F}_2=\sigma(\mathscr{H})$，由(2.10.41)立即得出$f(\omega^*|\omega;\mathscr{F}_2)$具有性质①。由(2.10.42)看出性质②对除去$N$中的$\omega$外皆成立。由于$P(N)=0$，所以$N$中的$\omega$可以不考虑它；如有必要，则通过“任意赋值”使它具有性质②。

由(2.10.32)，(2.10.34)和(2.10.39)得出：对任意的$A\in\mathscr{F}$，$f(\omega^*|\omega;\mathscr{F}_2)$满足(2.10.43)。得证$f(\omega^*|\omega;\mathscr{F}_2)$具有性质③。 **证毕**

本小段小结：① 离散型概率空间中确定各种类型条件概率的点函数——条件分布律，条件密度，条件密度—概率和条件密度—分布律皆有简单的、明确的表达式。它们为一般概率空间的研究提供参考。

② 由于定理2.8.1的原因，研究离散型概率空间时人们把分布律，而不是概率测度作为研究的出发点。我们也是这样做的。

③ 预先把离散型空间标准化也是克服分布律不惟一的一种方法。因为标准化后的概率空间保持概率测度$P(\mathscr{F})$不变。并且出现十分完美的情形：$P(\mathscr{F})$与分布律$f(\omega)$相互惟一确定；条件概率$P(\mathscr{F}|B)$与条件分布律$f(\omega|B)$相互惟一确定；条件概率$P(A|\mathscr{F}_2)$与条件密度$p(A|\mathscr{F}_2)(\omega)$相互惟一确定；条件密度—概率$p(\mathscr{F}|\omega;\mathscr{F}_2)$与条件密度—分布律$f(\omega^*|\omega;\mathscr{F}_2)$相互惟一确定。

2.10.4 *n*维柯氏概率空间上的点函数

n维柯氏概率空间中各种类型的条件概率皆存在确定它们的点函数。这时的点函数是欧氏空间的普通函数，2.10.1和2.10.2段中关于测度的抽象积分可以转化为人们比较熟悉的勒贝格—斯蒂尔吉斯积分(以下简称为L-S积分)。在相当宽广的条件下L-S积分是级数或黎曼(B. Riemann)积分。因此，如果愿意放弃一般

性，可以在初等微积分基础上介绍条件概率的理论。

在本段中$(\mathbf{R}^n,\mathscr{B}^n,P)$是$n$维柯氏概率空间，$F(x_1,x_2,\cdots,x_n)$是$P(\mathscr{B}^n)$的分布函数，它存在且惟一。

(1) 条件分布函数$F(x_1,x_2,\cdots,x_n|B)$

由引理2.9.1的推论知，条件概率$P(\mathscr{B}^n|B)$是$(\mathbf{R}^n,\mathscr{B}^n)$上的概率测度。由定理2.8.3得出，$P(\mathscr{B}^n|B)$与它的分布函数$F^*(x_1,x_2,\cdots,x_n)$相互惟一确定。为了反映$F^*$依赖于$B$，按习惯把$F^*$改为$F(x_1,x_2,\cdots,x_n|B)$。我们有

$$F(x_1,x_2,\cdots,x_n|B)=P\Big(\prod_{i=1}^{n}(-\infty,x_i]\,|B\Big)=\frac{P(B\Pi(x))}{P(B)}$$
$$=\frac{1}{P(B)}\int_{B\Pi(x)}\mathrm{d}_{x_1}\mathrm{d}_{x_2}\cdots\mathrm{d}_{x_n}F(x_1,x_2,\cdots,x_n) \tag{2.10.44}$$

其中$\Pi(x)=\prod_{i=1}^{n}(-\infty,x_i]$。此式表明$F(x_1,x_2,\cdots,x_n|B)$由分布律$F$惟一确定。

定义 2.10.5 称条件概率$P(\mathscr{B}^n|B)$的分布函数

$$F(x_1,x_2,\cdots,x_n|B)=P\Big(\prod_{i=1}^{n}(-\infty,x_i]\,|B\Big),$$
$$(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n \tag{2.10.45}$$

为分布函数F关于事件B的条件分布函数，简称为关于B的条件分布函数。

下面是用关于B的条件分布函数确定关于B的条件概率（测度）的基本结论。

定理 2.10.9 柯氏概率空间$(\mathbf{R}^n,\mathscr{B}^n,P)$中任意的条件概率$P(\mathscr{B}_1|B)$由关于$B$的条件分布函数$F(x_1,x_2,\cdots,x_n|B)$用公式

$$P(A|B)=\int_A\mathrm{d}_{x_1}\mathrm{d}_{x_2}\cdots\mathrm{d}_{x_n}F(x_1,x_2,\cdots,x_n|B)\quad(A\in\mathscr{B}_1) \tag{2.10.46}$$

确定。

证明 由定理 2.8.3 的推论知，条件概率 $P(\mathscr{B}^n|B)$由 $F(x_1,x_2,\cdots,x_n|B)$用公式

$$P(A|B)=\int_A \mathrm{d}_{x_1}\mathrm{d}_{x_2}\cdots d_{x_n}F(x_1,x_2,\cdots,x_n|B),(A\in\mathscr{B}^n)$$

确定。注意到 $P(\mathscr{B}_1|B)=P(\mathscr{B}^n|B)\upharpoonright\mathscr{B}_1$，故上式包含(2.10.46)。 **证毕。**

(2) 条件密度 $p(A|\mathscr{B}_2)(y_1,y_2,\cdots,y_n)$

应用 L-S 积分，定理 2.10.1 在柯氏概率空间中可以叙述为

定理 2.10.10 给定$(\mathbf{R}^n,\mathscr{B}^n,P)$和子试验$(\mathbf{R}^n,\mathscr{B}_2)$。那么对任意的 $A\in\mathscr{B}^n$，条件概率 $P(A|\mathscr{B}_2)$存在条件密度 $p(A|\mathscr{B}_2)(y_1,y_2,\cdots,y_n)$，$(y_1,y_2,\cdots,y_n)\in\mathbf{R}^n$。即 $p(A|\mathscr{B}_2)$具有性质：

① 它是 n 元 $\mathscr{B}_2$ 可测函数；

② 对任意的 $B\in\mathscr{B}_2^*$ 成立；

$$P(A|B)$$
$$=\frac{1}{P(B)}\int_B p(A|\mathscr{B}_2)(y_1,y_2,\cdots,y_n)\mathrm{d}_{y_1}\mathrm{d}_{y_2}\cdots\mathrm{d}_{y_n}F(y_1,y_2,\cdots,y_n)\tag{2.10.47}$$

并且性质①和②惟一地确定 $p(A|\mathscr{B}_2)$。其中 $F(y_1,y_2,\cdots,y_n)$是 $P(\mathscr{B}^n)$的分布函数。

在 2.10.1 中关于条件密度的说明皆适用于 $p(A|\mathscr{B}_2)$。

(3) 条件密度—概率 $p(A|y_1,y_2,\cdots,y_n;\mathscr{B}_2)$

应用定理 2.10.4，可以把定理 2.10.10 加强成

定理 2.10.11 给定$(\mathbf{R}^n,\mathscr{B}^n,P)$。那么对任意的子试验$(\mathbf{R}^n,\mathscr{B}_1)$和$(\mathbf{R}^n,\mathscr{B}_2)$，条件概率 $P(\mathscr{B}_1|\mathscr{B}_2)$存在条件密度—概率 $p(A|y_1,y_2,\cdots,y_n;\mathscr{B}_2)$，$(A,(y_1,y_2,\cdots,y_n))\in\mathscr{B}_1\times\mathbf{R}^n$。即 $p(A|y_1,y_2,\cdots,y_n;\mathscr{B}_2)$具有三个性质：

① 对固定的 A，它关于$(y_1,y_2,\cdots,y_n)$是 $\mathscr{B}_2$ 可测函数；

② 对固定的$(y_1,y_2,\cdots,y_n)$，它关于 A 是$(\mathbf{R}^n,\mathscr{B}_1)$上的概率测度；

③ 对任意的 $A\in\mathscr{B}_1$，$B\in\mathscr{B}_2^*$ 成立

$$P(A|B)=\frac{1}{P(B)}\int_B p(A|y_1,y_2,\cdots,y_n;\mathscr{B}_2)\mathrm{d}_{y_1}\mathrm{d}_{y_2}\cdots\mathrm{d}_{y_n}\times F(y_1,y_2,\cdots,y_n) \tag{2.10.48}$$

其中 $F(y_1,y_2,\cdots,y_n)$是 $P(\mathscr{B}^n)$的分布函数。

(4) 条件密度—分布函数 $F(x_1,x_2,\cdots,x_n|y_1,y_2,\cdots,y_n;\mathscr{B}_2)$

在定理 2.10.11 中取 $\mathscr{B}_1=\mathscr{B}^n$ 得到$(\mathbf{R}^n,\mathscr{B}^n)$上的概率测度 $p(\mathscr{B}^n|y_1,y_2,\cdots,y_n;\mathscr{B}_2)$。若用 $F(x_1,x_2,\cdots,x_n|y_1,y_2,\cdots,y_n;\mathscr{B}_2)$表示它的分布函数，并注意到$(y_1,y_2,\cdots,y_n)$的任意性，便可产生

定义 2.10.6 用 $p(\mathscr{B}^n|\mathbf{R}^n;\mathscr{B}_2)$表示 $\mathscr{B}^n$ 关于 $\mathscr{B}_2$ 的条件密度—概率，则称

$$F(x_1,x_2,\cdots,x_n|y_1,y_2,\cdots,y_n;\mathscr{B}_2)$$

$$=p\left(\prod_{i=1}^{n}(-\infty,x_i]|y_1,y_2,\cdots,y_n;\mathscr{B}_2\right) \tag{2.10.49}$$

为分布函数 F 关于$(\mathbf{R}^n,\mathscr{B}_2)$的条件密度—分布函数。简称为关于 $\mathscr{B}_2$ 的条件密度—分布函数。

定理 2.10.12 关于 $\mathscr{B}_2$ 的条件密度—分布函数 $F(\mathbf{R}^n|\mathbf{R}^n;\mathscr{B}_2)$具有下列性质：

① 对固定的$(x_1,x_2,\cdots,x_n)$，它关于$(y_1,y_2,\cdots,y_n)$是 $\mathscr{B}_2$ 可测度函数；

② 对固定的$(y_1,y_2,\cdots,y_n)$，它关于$(x_1,x_2,\cdots,x_n)$是 n 元分布函数；

③ 对任意的子试验$(\mathbf{R}^n,\mathscr{B}_1)$，它用公式

$$
\begin{aligned}
&p(A \mid y_1, y_2, \cdots, y_n; \mathscr{B}_2) \\
&= \int_A \mathrm{d}_{x_1} \mathrm{d}_{x_2} \cdots \mathrm{d}_{x_n} F(x_1, x_2, \cdots x_n \mid y_1, y_2, \cdots, y_n; \mathscr{B}_2) \quad (A \in \mathscr{B}_1)
\end{aligned}
\tag{2.10.50}
$$

确定 $\mathscr{B}_1$ 关于 $\mathscr{B}_2$ 的条件密度一概率 $p(\mathscr{B}_1 \mid \mathbf{R}^n, \mathscr{B}_2)$。

证明 性质①和②由(2.10.49)立即得出。由定理 2.8.3 的推论得出(2.10.50)对任意的 $A \in \mathscr{B}^n$ 成立,故 $F(\mathbf{R}^n \mid \mathbf{R}^n; \mathscr{B}_2)$ 具有性质③。 **证毕**

2.10.5 注记

在(2.10.1)中令 $B=\Omega$ 得到

$$
P(A) = \int_\Omega p(A \mid \mathscr{F}_2)(\omega) P(\mathrm{d}\omega) \tag{2.10.51}
$$

它是一般形式的全概率公式。

事实上,取 $\mathscr{F}_2$ 为定理 2.9.9 中的 $\sigma(\mathscr{H})$ 时(2.10.51)成为(2.9.43)(参看(2.10.35))。(2.10.51)可以解释为用 Ω 的最细划分代替划分 $\mathscr{H}$ 后(2.9.43)的推广式。

2.11 第Ⅵ组公理:概率建模公理组

2.1 节已指明,应用第Ⅰ～Ⅴ组公理可以推演出概率论的理论体系。因此,这五组公理组成了概率论的形式化公理系统。

概率论是应用性很强的一门学科。正如 2.4 节第(4)段所述,概率论的应用要求人们把所关心的随机现象抽象为理论体系中的研究对象(如实现(2.7.4)的建模工作,3.1 节中的工作等),然后对研究对象进行理论推演,得到各种各样的结论,并用这些结论表达出所关心随机现象蕴含的统计知识和统计规律。

为了规范建模工作(2.7.4),2.4 节讨论了建立随机试验

$(\Omega,\mathscr{F})$的工作，本节讨论建立概率测度 $P(\mathscr{F})$的工作，并引进一组新公理。这组公理把久经考验的三条原理(等可能性原理，独立实验原理和概率的频率解释)抽象为六个公理。应用它们能够建立一些常见的概率空间——古典型、离散型、几何型、独立乘积型和柯氏型概率空间。

本组公理是开放性的。期待用更好的、更行之有效的公理充实它、修正它。本组公理的直观背景包含在 1.3 节中。

2.11.1 概率建模公理组

设$(\Omega,\mathscr{F})$是随机试验，$A,A_i\in\mathscr{F}(i=1,2,\cdots,n;n\geqslant 2)$，并且 $\mathscr{A}=\{A_1,A_2,\cdots,A_n\}$是 A 的一个划分。

公理 Ⅵ₁(等可能性公理)　假定 $P(A)$已知，A 的任何真子集皆未赋概。如果认定诸事件 $A_1,A_2,\cdots,A_n$ 有同样的出现可能性，则赋予概率

$$P(A_1)=P(A_2)=\cdots=P(A_n)=\frac{1}{n}P(A)\quad(2.11.1)$$

称为用划分 $\mathscr{A}$ 对 A 进行等可能性处理。

公理 Ⅵ₂(赋值公理)　假定 $P(A)$已知，A 的任何真子集皆未赋概。如果有理由(客观的或主观的)认定 A_i 有概率值 $P(A_i)$ $(i=1,2,\cdots,n)$，并使得

$$P(A_1)+P(A_2)+\cdots+P(A_n)=P(A)\quad(2.11.2)$$

则认为“理由”是合理的、可接受的。称这种赋概为用 $P(A_i)(A_i\in\mathscr{A})$对 A 进行赋值处理。

今后，称 $\mathscr{A}$ 中元素为使用划分块，简称为使用块。显然，等可能性处理是一种特殊的赋值处理。

公理 Ⅵ₃(顺序公理)　当赋值公理多次使用时必须进行分层。第一层是对某事件 A 进行；第 $k+1$ 层只能对 k 层的某些使用块(假定为 $D_{k_1},D_{k_2},\cdots,D_{k_r}$)进行赋值处理$(k=1,2,3,\cdots)$。并且存在

正整数 M 使得 $k_r \leqslant M(k=1,2,3,\cdots)$ 和成立

$$\lim_{k\to\infty} P\left(\bigcup_{i=1}^{k_r} D_{k_i}\right) = 0 \qquad (2.11.3)$$

称使用块是终极的，如果它是第 k 层的使用块，第 $k+1$ 块不对它进行赋值处理。此公理中的 M 保证终极使用块的总数或为有限个，或为可列个。(2.11.3)保证使用块所赋概率遵守公理Ⅳ$_5$。

公理Ⅵ$_4$(几何赋概公理)　对 n 维波莱尔试验$(\mathbf{R}^n,\mathscr{B}^n)$使用顺序公理时不受正整数 M 和(2.11.3)的限制，但要求 $k\to\infty$ 时第 k 层的每个使用块的概率趋于零。

公理Ⅵ$_5$(独立实验公理)　假定实验 $\mathscr{E}_j$ 产生概率空间$(\Omega_j,\mathscr{F}_j,P_j)$($j\in J$,$J$ 为任意指标集)。如果对任意的 $j\in J$，实验 $\mathscr{E}_j$ 的结果(指 $\mathscr{F}_j$ 中事件及其概率)不受其他实验的影响，则联合实验 $\mathscr{E}^*$ 产生的概率空间是$\left(\prod_{j\in J}\Omega_j,\prod_{j\in J}\mathscr{F}_j,\prod_{j\in J}P_j\right)$。

特别，假定 $\mathscr{E}$ 产生$(\Omega,\mathscr{F},P)$。用 $\mathscr{E}^n$ 和 $\mathscr{E}^\infty$ 分别表示把 $\mathscr{E}$ 独立地、重复地进行 n 次或可列次产生的新实验。那么 $\mathscr{E}^n$ 和 $\mathscr{E}^\infty$ 分别产生独立重复概率空间$(\Omega^n,\mathscr{F}^n,P^n)$和$(\Omega^\infty,\mathscr{F}^\infty,P^\infty)$。

公理Ⅵ$_6$(频率公理)　设实验 $\mathscr{E}$ 产生概率空间$(\Omega,\mathscr{F},P)$。如果把 $\mathscr{E}$ 独立地、重复地进行可列次，那么对任意的 $A\in\mathscr{F}$，在$(\Omega^\infty,\mathscr{F}^\infty,P^\infty)$中成立

$$P^\infty\left\{\omega \,\middle|\, \lim_{n\to\infty}\frac{\nu_{nA}(\omega)}{n}=P(A)\right\}=1^{①} \qquad (2.11.4)$$

这里 $\omega=(\omega_1,\omega_2,\cdots,\omega_n,\cdots)\in\Omega^\infty$，$\nu_{nA}(\omega)$表示 ω 的前 n 个坐标中使得 $\omega_i\in A(i=1,2,\cdots,n)$的个数。

直观上，$\nu_{nA}(\omega)$表示前 n 次进行实验 $\mathscr{E}$ 时事件 A 出现的次数。因此称$\frac{\nu_{nA}(\omega)}{n}$为事件 A 在样本点 ω 上的频率。公理Ⅵ$_6$ 表明 $n\to\infty$

① 习惯上，这个式子又可写为$\lim_{n\to\infty}\frac{\nu_{nA}(\omega)}{n}=P(A)$　(a.s. ω)

时事件 A 的频率的极限存在且等于 A 的概率。定理 2.7.8 保证公理 Ⅵ$_6$ 与前五组公理相容。同样，定理 2.9.6 的推论保证公理 Ⅵ$_5$ 也有这样的相容性。

2.11.2 古典型概率空间建模

定理 2.11.1 设$(\Omega,\mathscr{F})$是 n 型标准试验，$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$。如果人们接受用划分$\{\{\omega_i\}\mid i=1,2,\cdots,n\}$对 Ω 进行等可能性处理，那么公理 Ⅵ$_1$ 在$(\Omega,\mathscr{F})$上赋予古典型概率测度 $P(\mathscr{F})$，并得到古典型概率空间$(\Omega,\mathscr{F},P)$。

证明 使用公理 Ⅵ$_1$ 后得到(2.7.10)。应用引理 2.7.2 及其推论即得定理结论。 **证毕**

现实中的 n 型标准试验是否能用最细划分对 Ω 进行等可能性处理视情况而定，并最终要经受实践的检验。通常，人们能够接受对 2.4 节中例 1，例 5 和例 8 的等可能性处理。根据孟德尔遗传学理论，对 2.4 节例 3 能进行等可能性处理，对例 4 则不能。

2.11.3 离散型概率空间建模

给定随机试验$(\Omega,\mathscr{F})$。假定对它已经使用顺序公理且第一层处理的事件是必然事件 Ω。用 $H_1,H_2,\cdots,H_n,\cdots$ 表示全体终极使用块，且令 $H_0=\Omega\setminus\bigcup\limits_{i\geqslant1}H_i$（它可能是空集）。因此

$$\mathscr{H}=\{H_0,H_1,H_2,\cdots\cdots\}\tag{2.11.5}$$

是 Ω 的一个划分。用 f_i 表示顺序公理赋予事件 H_i 的概率值$(i\geqslant1)$，由(2.11.3)推出 $\sum\limits_{i\geqslant1}f_i=P(\Omega)=1$。即是说，顺序公理赋予离散型标准试验$(\mathscr{H},\sigma(\mathscr{H}))$上一个概率分布律

$$f:\begin{pmatrix}H_1 & H_2 & \cdots & H_n & \cdots & H_0\\ f_1 & f_2 & \cdots & f_n & \cdots & 0\end{pmatrix}\tag{2.11.6}$$

应用定理 2.8.1 得出这个分布律在$(\mathscr{H},\sigma(\mathscr{H}))$上惟一地确定一

个概率测度 $P(\sigma(\mathscr{H}))$。于是证明了

引理 2.11.1 给定随机试验$(\Omega,\mathscr{F})$。在其上使用顺序公理后得到子试验$(\Omega,\sigma(\mathscr{H}))$上的概率测度 $P(\sigma(\mathscr{H}))$和离散型标准概率空间$(\mathscr{H},\sigma(\mathscr{H}),P)$。

例 1 假定硬币无破损,桌面光滑,投掷时无倾向性。试建立 2.4 节例 9 中试验$(\Omega_9,\mathscr{F}_9)$上的概率测度。

解 在假定的条件下人们接受用划分$\{\{\omega_1\},\{\omega_2,\omega_3,\cdots,\omega_\infty\}\}$对 Ω_9 进行等可能性处理。于是

$$P(\{\omega_1\}) = P(A_1) = \frac{1}{2} \tag{2.11.7}$$

这里 $A_1=\{\omega_2,\omega_3,\cdots,\omega_\infty\}$。第二层处理只对 A_1 进行。人们接受用划分$\{\{\omega_2\},\{\omega_3,\omega_4,\cdots,\omega_\infty\}\}$对 A_1 进行等可能性处理。于是

$$P(\{\omega_2\}) = P(A_2) = \frac{1}{2}P(A_1) = \frac{1}{4} \tag{2.11.8}$$

这里 $A_2=\{\omega_3,\omega_4,\cdots,\omega_\infty\}$。如此继续,第 n 层处理只对 $A_{n-1}=\{\omega_n,\omega_{n+1},\cdots,\omega_\infty\}$进行,且接受用划分$\{\{\omega_n\},\{\omega_{n+1},\omega_{n+2},\cdots,\omega_\infty\}\}$对 A_{n-1}进行等可能性处理。于是

$$P(\{\omega_n\}) = P(A_n) = \frac{1}{2}P(A_{n-1}) = \frac{1}{2^n} \tag{2.11.9}$$

这里 $A_n=\{\omega_{n+1},\omega_{n+2},\cdots,\omega_\infty\}$。由归纳法,能够对任意正整数层进行处理。

对$(\Omega_9,\mathscr{F}_9)$使用顺序公理后得到终极使用块$\{\omega_i\}(i=1,2,3,\cdots)$(注意,它不包括$\{\omega_\infty\}$)。由引理 2.11.1 得出顺序公理在$(\Omega_9,\mathscr{F}_9)$上赋予分布律

$$\begin{pmatrix} \omega_1 & \omega_2 & \cdots & \omega_n & \cdots & \omega_\infty \\ \frac{1}{2} & \frac{1}{4} & \cdots & \frac{1}{2^n} & \cdots & 0 \end{pmatrix} \tag{2.11.10}$$

用 $P(\mathscr{F}_9)$表示它产生的概率测度。最终,我们用顺序公理建立了 2.4 节例 9 中实验产生的概率空间$(\Omega_9,\mathscr{F}_9,P)$。

定理 2.11.2 设$(\Omega,\mathscr{F},P)$是任意的离散型概率空间。假定$(\Omega,\mathscr{F})$已知，那么能够用顺序公理在其上赋予概率测度$P(\mathscr{F})$。

证明 不失一般性，设

$$\Omega = \{\omega_1,\omega_2,\cdots,\omega_n,\cdots\} \tag{2.11.11}$$

$P(\mathscr{F})$的分布律为

$$\begin{pmatrix} \omega_1 & \omega_2 & \cdots & \omega_n & \cdots \\ f_1 & f_2 & \cdots & f_n & \cdots \end{pmatrix} \tag{2.11.12}$$

首先假定$(\Omega,\mathscr{F})$是标准的。现在对它使用顺序公理。第一层处理是用划分$\{\{\omega_1\},A_1\}$和

$$P(\{\omega_1\}) = f_1 \quad P(A_1) = 1 - f_1 \tag{2.11.13}$$

对Ω进行赋值处理，这里$A_1=\{\omega_2,\omega_3,\cdots\}$。

第n层处理是用$A_{n-1}=\{\omega_n,\omega_{n+1},\cdots\}$的划分$\{\{\omega_n\},A_n\}$和

$$P(\{\omega_n\}) = f_n, \quad P(A_n) = 1 - \sum_{i=1}^{n} f_i \tag{2.11.14}$$

对A_{n-1}进行赋值处理，这里$A_n=\{\omega_{n+1},\omega_{n+2},\cdots\}$。

应用归纳法能够对任意正整数层进行处理。对$(\Omega,\mathscr{F})$使用顺序公理的结果是获得分布律(2.11.12)，从而得到概率测度$P(\mathscr{F})$。

随后假定$(\Omega,\mathscr{F})$是非标准的。那么应用定理2.5.3(或定理2.5.6)把它化为标准的，再用已证的结论即完成证明。 **证毕**

2.11.4 几何型概率空间建模

设$(D,\mathscr{B}^n(D))$是D上的波莱尔试验，μ是$(\mathbf{R}^n,\mathscr{B}^n)$上的勒贝格测度。假定$0<\mu(D)<+\infty$。

由勒贝格测度的理论得知，存在D的子集D_0和$D_1=D\backslash D_0$使得$\mu(D_0)=\mu(D_1)$。如果认定D_0和D_1有同样的出现可能性，那么人们接受用划分$\{D_0,D_1\}$对必然事件D进行等可能性处理。于是

$$P(D_0)=P(D_1)=\frac{1}{2}=\frac{\mu(D_{i_1})}{\mu(D)} \quad (i_1=0 \text{ 或 } 1) \qquad (2.11.15)$$

这是顺序公理中的第一层处理。

第二层处理同时对两个使用块 $D_{i_1}(i_1=0$ 或 $1)$进行。每个使用块进行类似于 D 的等可能性处理，得到四个使用块 $D_{i_1i_2}(i_1,i_2=0,1)$和

$$P(D_{i_1i_2}) = \frac{1}{2}P(D_i) = \frac{1}{4} = \frac{\mu(D_{i_1i_2})}{\mu(D)} \quad (i_1,i_2=0,1) \qquad (2.11.16)$$

假定前 $n-1$ 层处理完毕，得到使用块 $D_{i_1i_2\cdots i_{n-1}}(i_1,i_2,\cdots,i_{n-1}=0,1)$。公理 Ⅵ$_4$ 保证第 n 层可以对此 2^{n-1}个使用块同时进行类似于 D 的等可能性处理。实施细则为，先获得 $D_{i_1i_2\cdots i_{n-1}}$ 的划分 $\{D_{i_1i_2\cdots i_{n-1}0},D_{i_1i_2,\cdots i_n,1}\}$且它们有相同的勒贝格测度。然后认定它们有同样的出现可能性，于是人们接受用这个划分对 $D_{i_1i_2\cdots i_{n-1}}$进行等可能性处理且得到

$$P(D_{i_1i_2\cdots i_n}) = \frac{1}{2}P(D_{i_1i_2\cdots i_{n-1}}) = \frac{1}{2^n} = \frac{\mu(D_{i_1i_2\cdots i_n})}{\mu(D)} \quad (i_1,i_2,\cdots,i_n=0,1) \qquad (2.11.17)$$

如此继续，归纳法保证可以对任意正整数层进行处理。

用 $\mathscr{D}$ 表示应用顺序公理后的所有使用块组成的集合。它是

$$\mathscr{D} = \{D_{i_1i_2\cdots i_n} | n \geqslant 1, i_1,i_2,\cdots,i_n=0,1\} \qquad (2.11.18)$$

众所周知，对适当选择的使用块（例如，当 D 是有界集时只须要求 $n\to\infty$时 n 层使用块的直径趋于零）成立 $\mathscr{B}^n(D)=\sigma(\mathscr{D})$和

$$P(B) = \frac{\mu(B)}{\mu(D)} \quad (B \in \mathscr{B}^n(D)) \qquad (2.11.19)$$

于是证明了

定理 2.11.3 设$(D,\mathscr{B}^n(D))$是D上的波莱尔试验且$0<\mu(D)<+\infty$,那么应用公理Ⅵ$_1$,Ⅵ$_3$和Ⅵ$_4$可以在其上赋予几何概率测度$P(\mathscr{B}^n(D))$和得到几何型概率空间$(D,\mathscr{B}^n(D),P)$。

2.11.5 n维柯氏概率空间建模

定理 2.11.4 设$(\mathbf{R}^n,B^n,P)$是任意的n维柯氏概率空间。那么能够应用公理Ⅵ$_1$～Ⅵ$_4$在$(\mathbf{R}^n,\mathscr{B}^n)$上赋予概率测度$P(\mathscr{B}^n)$。

证明 只证$n=1$的情形。$n\neq1$时证明类似,只是记号烦琐。

用$F(x)$表示$P(\mathscr{B})$的分布函数。首先假定$F(x)$是连续函数。这时存在实数a_1使得$F(a_1)=\frac{1}{2}$。如果认定$D_0=(-\infty,a_1]$与$D_1=(a_1,+\infty)$有同样的出现可能性,那么人们能够接受用划分$\{D_0,D_1\}$对必然事件$\mathbf{R}$进行等可能性处理,得到

$$P(D_{i_1})=\frac{1}{2}=F(\beta_{i_1})-F(\alpha_{i_1})\quad(i_1=0,1)\tag{2.11.20}$$

其中$(\alpha_{i_1},\beta_{i_1}]$代表使用块$D_{i_1}$,并且允许$\alpha_{i_1},\beta_{i_1}$取$-\infty$和$+\infty(i_1=0,1)$。这是第一层处理。

第二层处理将对两个使用块D_0和D_1同时进行。由$F(x)$的连续性推出,存在实数a_{10}和a_{11}使得$F(a_{10})=\frac{1}{4}$,$F(a_{11})=\frac{3}{4}$。令

$$D_{00}=(-\infty,a_{10}],D_{01}=(a_{10},a_1],$$
$$D_{10}=(a_1,a_{11}],D_{11}=(a_{11},+\infty)$$

如果认定D_{i_10}和$D_{i_11}(i_1=0$或$1)$有同样的出现可能性,那么人们接受用划分$\{D_{i_10},D_{i_11}\}$对D_{i_1}进行等可能性处理,得到

$$\begin{aligned}P(D_{i_1i_2})&=\frac{1}{2}P(D_{i_1})=\frac{1}{4}\\&=F(\beta_{i_1i_2})-F(\alpha_{i_1i_2})\ (i_1,i_2=0,1)\end{aligned}\tag{2.11.21}$$

其中$(\alpha_{i_1i_2},\beta_{i_1i_2}]$是使用块$D_{i_1i_2}(i_1,i_2=0,1)$。

假定前$n-1$层处理进行完毕。第$n-1$层的使用块是

$$D_{i_1i_2\cdots i_{n-1}}=(\alpha_{i_1i_2\cdots i_{n-1}},\beta_{i_1i_2\cdots i_{n-1}}],(i_1,i_2,\cdots,i_{n-1}=0,1) \tag{2.11.22}$$

公理Ⅵ$_4$保证第n层处理可以同时对这些(共2^{n-1}个)使用块进行处理。用$(\alpha,\beta]$作为(2.11.22)中区间的代表(省略下标)。由$F(x)$的连续性得知,在(α,β)中存在点γ使得$F(\gamma)=F(\alpha)+\frac{1}{2^n}$。如果认定$(\alpha,\gamma]$与$(\gamma,\beta]$有同样的出现可能性,那么人们接受用划分$\{(\alpha,\gamma],(\gamma,\beta]\}$对$(\alpha,\beta]$进行等可能性处理。于是得到$P((\alpha,\gamma])=P((\gamma,\beta])=\frac{1}{2}P((\alpha,\beta])=\frac{1}{2^n}$。恢复下标后得到

$$P(D_{i_1i_2\cdots i_n})=\frac{1}{2}P(D_{i_1i_2\cdots i_{n-1}})=\frac{1}{2^n}=F(\beta_{i_1i_2\cdots i_n})-F(\alpha_{i_1i_2\cdots i_n})$$
$$(i_1,i_2,\cdots,i_n=0,1) \tag{2.11.23}$$

其中$(\alpha_{i_1i_2\cdots i_n},\beta_{i_1i_2\cdots i_n}]$是第$n$层处理的使用块$D_{i_1i_2\cdots i_n}$。

公理Ⅵ$_3$和Ⅵ$_4$保证可以进行任意正整数层处理,使用公理Ⅵ$_3$完毕后把所有使用块组成集合$\mathscr{D}$,即

$$\mathscr{D}=\{D_{i_1i_2\cdots i_n}\mid n\geqslant 1,i_1,i_2,\cdots,i_n=0,1\} \tag{2.11.24}$$

众所周知,$\mathscr{B}=\sigma(\mathscr{D})$,上述赋予的数值$P(\mathscr{D})$惟一地扩张成$\mathscr{B}$上的概率测度$P(\mathscr{B})$。下证$P(\mathscr{B})$的分布函数是$F(x)$,即证明

$$F(x)=P((-\infty,x]) \tag{2.11.25}$$

事实上,若x是使用块的端点,等式是等可能性处理与概率有限可加性的结果。对一般的x,这个等式可由$F(x)$的连续性与概率测度的上、下连续性推出。得证$P(\mathscr{B})$是定理中给定的概率测度。

随后假定$F(x)$是一般的分布函数。应用勒贝格分解定理得

$$F(x)=\alpha_1F_1(x)+\alpha_2F_2(x) \tag{2.11.26}$$

其中 $F_1(x)$ 是离散型分布函数[①]，$F_2(x)$ 是连续的分布函数，α_1，$\alpha_2 \geqslant 0$ 且 $\alpha_1+\alpha_2=1$。

于是，应用定理 2.11.2 可以在 $(\mathbf{R},\mathscr{B})$ 上赋予分布函数 $F_1(x)$，应用已证的结论可以赋予分布函数 $F_2(x)$。从而可以赋予分布函数 $F(x)$，并用它产生概率测度 $P(\mathscr{B})$。 **证毕**

2.11.6 独立乘积概率空间建模

公理Ⅵ$_5$ 给出独立乘积概率空间的建模方法，它在应用中已经广泛使用。下述定理解释古典型概率空间在应用中大量出现的原因。

定理 2.11.5 设实验 $\mathscr{E}_i$ 产生古典型概率空间 $(\Omega^i,\mathscr{F}^i,P^i)$。假定诸实验 $\mathscr{E}_i(i=1,2,\cdots,n)$ 相互无关联（即遵守公理Ⅵ$_5$ 的条件）。用 $\mathscr{E}^*$ 表示 $\mathscr{E}_1,\mathscr{E}_2,\cdots,\mathscr{E}_n$ 的联合实验，那么 $\mathscr{E}^*$ 产生的独立乘积概率空间 $\left(\prod_{i=1}^n\Omega^i,\prod_{i=1}^n\mathscr{F}^i,\prod_{i=1}^n P^i\right)$ 是古典型的。

证明 不妨假定 Ω^i 含有 n_i 个元素。在这个假定下可以写出

$$\prod_{i=1}^n\Omega^i=\{(\omega_1,\omega_2,\cdots,\omega_n)\,|\,\omega_i\in\Omega^i,i=1,2,\cdots n\} \quad (2.11.27)$$

$$\prod_{i=1}^n\mathscr{F}^i=\{A_1\times A_2\times\cdots\times A_n\,|\,A_i\in\mathscr{F}^i,i=1,2,\cdots n\}$$

$$(2.11.28)$$

并且 $\prod_{i=1}^n\Omega^i$ 含有 $n_1n_2\cdots n_n$ 个样本点。把乘积概率测度 $\prod_{i=1}^n P^i$ 简记为 P，我们有

$$P(A_1\times A_2\times\cdots\times A_n)=P^1(A_1)P^2(A_2)\cdots P^n(A_n)$$

① 用 $\Omega=\{x_1,x_2,\cdots\}$ 表示 $F(x)$ 的间断点的集合，众所周知，$F_1(x)$ 与 Ω 上的某分布律 $f(x_n)(n\geqslant 1)$ 相互惟一确定。

$$= \prod_{i=1}^{n} \frac{A_i \text{ 含 } \Omega^i \text{ 中样本点的个数}}{n_i}$$

$$= \frac{A_1 \times A_2 \times \cdots \times A_n \text{ 含 } \Omega^1 \times \Omega^2 \times \cdots \times \Omega^n \text{ 中样本点的个数}}{n_1 n_2 \cdots n_n}$$

(2.11.29)

得证$\left(\prod_{i=1}^{n}\Omega^i, \prod_{i=1}^{n}\mathscr{F}^i, \prod_{i=1}^{n}P^i\right)$是古典型概率空间。 **证毕**

2.11.7 频率公理的重要性

现实中人们无法把一个实验$\mathscr{E}$无限次重复地进行。因此公理Ⅵ$_6$中的等式

$$\lim_{n\to\infty} \frac{\nu_{nA}(\omega)}{n} = P(A) \quad (\text{a.s.}\ \omega) \qquad (2.11.30)$$

不能用于求$P(A)$的精确值,只能确定它的近似值。即便如此,人们仍然可以按1.3节第(5)段的方法进行建模,孟德尔的工作是最光辉的建模范例之一。

频率公理在概率论中的重要性在于:

(1) (2.11.30)是理解抽象概念"概率"的重要工具。

(2) (2.11.30)是检验概率论理论体系是否有效的重要工具。

(3) (2.11.30)是数理统计的基础。在这个基础上发展出的统计方法成为发现、建立和检验其他科学理论或科学规律的一种重要的科学方法。

等可能性原理和概率的频率解释是支撑整个概率论大厦的两大支柱。因此自然公理系统不把它们视为前五组公理的推论,而把它们作为并列的公理,成为建模的依据。

第3章　随机变量简介

自然公理系统在引进因果空间后又引进了空间中的两类“图形”——随机试验和概率空间，以及研究它们的工具——联合试验和条件概率。于是，公理化后概率论转向研究因果空间中各种各样的“图形”，特别是各种各样的随机试验和概率空间。研究的基本任务可以归结为：

(1) 寻找应用中的随机试验，并研究不同试验(包括各种子试验)之间的关系。

(2) 根据应用问题的要求在试验$(\Omega,\mathscr{F})$上进行赋概，得到人们能接受的概率空间$(\Omega,\mathscr{F},P)$。在多数情况下只能获得$P(\mathscr{F})$的部分知识。

(3) 给定$(\Omega,\mathscr{F},P)$，寻找概率空间蕴含的统计知识，特别是其中的统计规律。

(4) 研究不同概率空间之间的关系，特别是关系中的统计知识和统计规律。

通常，数理统计侧重于研究基本任务(1)和(2)；概率论侧重于研究任务(3)和(4)[①]。如果我们把研究对象限于波莱尔试验和柯尔莫哥洛夫概率空间，则基本任务可以具体化为：

① 寻找应用中需要的波莱尔试验和它的各种子试验，并研究它们之间的关系。

② 根据应用问题的要求，寻找波莱尔试验上的分布函数(即

---

① “概率论”一词有广义的和狭义的理解方式。此处是按狭义的方式理解。本书中其他各处皆按广义的方式理解，即数理统计也是概率论的一部分。

在试验上赋概)，得到应用问题需要的柯尔莫哥洛夫概率空间。在多数场合中我们只能得到分布函数的部分知识。

③ 研究柯尔莫哥洛夫概率空间蕴含的统计知识，特别是其中的统计规律。

④ 研究不同柯尔莫哥洛夫概率空间之间的关系。特别是关系中的统计知识和统计规律。

随机变量或随机变量族是定义在$(\Omega,\mathscr{F},P)$上的一类特殊的点函数。由于在它的值域上存在柯氏型概率空间，因此我们能够用柯尔莫哥洛夫概率空间的统计知识获得$(\Omega,\mathscr{F},P)$的大量统计知识和统计规律。习惯上人们把这些知识和规律称为随机变量(或族)蕴含的统计知识和统计规律。即是说，在一定的意义下能够把上述的基本任务(1)～(4)转化为①～④，因此随机变量成为研究抽象概率空间的一种主要工具。

非常有趣的事情是，即使$(\Omega,\mathscr{F},P)$不能够具体写出，我们仍然可以借助于柯尔莫哥洛夫概率空间获得$(\Omega,\mathscr{F},P)$上随机变量(或族)的统计知识和统计规律。

另一方面，应用中所关心的随机现象皆来自于某个抽象的概率空间$(\Omega,\mathscr{F},P)$。在许多场合下，我们只能论证$(\Omega,\mathscr{F},P)$存在，但无法具体地写出它(例如 3.1 节中的天气预报实验)。在这种情况下如何研究所关心的随机现象呢?为此，人们引进刻画所关心的随机现象的数量化指标，并转为研究仅与指标有关的随机现象。例如，在掷骰子的实验中引进数量化指标"1,2,3,4,5,6"；在天气预报实验中引进温度、湿度……并且根据需要构造新的数量化指标。可以论证(参看 3.1 节)，数量化指标是随机变量；与指标有关的随机现象蕴含的统计知识和统计规律是随机变量的统计知识和统计规律。于是，随机变量提供了研究数量化指标的一种行之有效的方法，促进了现代科学，特别是社会科学定量化研究的进程。

鉴于上述的两个方面，随机变量成为概率论在公理化后主要

的研究对象。

本书的目的是建立自然公理系统，因此只能对随机变量进行非常简短的介绍。3.1 节论证随机变量成为概率论主要研究对象的必然性。3.2 节介绍随机变量概念和研究它的工具——分布函数和数字特征。3.3 和 3.4 节讨论随机变量之间的关系，介绍随机变量族概念和研究它们的工具——各种分布函数和数字特征，条件概率和条件数学期望。

3.1 随机变量的直观背景

3.1.1 直观背景：概率空间已知的情形

甲、乙、丙、丁四人博奕。庄家丁准备做 2.4 节例 5 中实验 $\mathscr{E}_5$：投掷一颗质料均匀的骰子一次。甲、乙、丙是赌客，他们和庄家约定的输赢规则分别为（假定 $a\neq 0$）

ξ	骰子掷出的点数	1	2	3	4	5	6	(3.1.1)
	甲赢钱数(元)	a	$-a$	a	$-a$	a	$-a$	
η	骰子掷出的点数	1	2	3	4	5	6	(3.1.2)
	乙赢钱数(元)	a	a	a	a	$-2a$	$-2a$	
ζ	骰子掷出的点数	1	2	3	4	5	6	
	丙赢钱数(元)	$-3a$	$-a$	$-3a$	$-a$	0	$2a$	(3.1.3)

若用 s 表示庄家丁赢钱的数目，则

$$s=-(\xi+\eta+\zeta) \tag{3.1.4}$$

s 也可以用表格表示为

s	骰子掷出的点数	1	2	3	4	5	6
	丁赢钱数(元)	a	a	a	a	a	a

(3.1.5)

四位博奕者关心的事情是"自己赢钱还是输钱?输赢多少钱?"姑且假定他们不关心其他的事情,那么他们只须关心骰子出现的点数和各自的输赢规则。让我们重点分析 ξ,即考察甲所关心的事情。

(1) 博奕存在的基础是概率空间$(\Omega_5,\mathscr{F}_5,P_5)$

博奕进行的前提是庄家做实验 $\mathscr{E}_5$。$\mathscr{E}_5$ 产生概率空间$(\Omega_5,\mathscr{F}_5,P_5)$,其中$(\Omega_5,\mathscr{F}_5)$是 2.4 节例 5 规定的随机试验,$P_5(\mathscr{F}_5)$是古典型概率测度。

(2) ξ 是定义在$(\Omega_5,\mathscr{F}_5,P_5)$上的点函数

表面上,甲赢钱数 ξ 是一个不确定的量。ξ 随机会而取值:掷出奇数点,赢 a 元钱;掷出偶数点,输 a 元钱。

样本空间的重大意义在于它的存在导致人们发现 ξ 是定义在 Ω_5 上的实值函数。事实上,(3.1.1)是它的表格式,它的解析式为

$$\xi(\omega)=\begin{cases} a, & \omega=1,3,5 \\ -a, & \omega=2,4,6 \end{cases} \tag{3.1.6}$$

(3) ξ 产生子试验$(\Omega_5,\sigma(\xi))$

常识表明,甲不关心 $\mathscr{F}_5$ 中的所有事件。他只关心 4 个事件,它们是

$$\{\varnothing,\{1,3,5\},\{2,4,6\},\Omega_5\} \tag{3.1.7}$$

其中 Ω_5 和$\varnothing$表达实验 $\mathscr{E}_5$ 是否进行;另外两个事件表达

$$\text{“赢 } a \text{ 元钱”}=\{1,3,5\}=\{\omega|\xi(\omega)=a\}\xlongequal{\text{def}}\{\xi=a\} \tag{3.1.8}$$

$$\text{“输 } a \text{ 元钱”}=\{2,4,6\}=\{\omega|\xi(\omega)=-a\}\xlongequal{\text{def}}\{\xi=-a\} \tag{3.1.9}$$

由此看出，一个事件有多种表示方法。左端的“文字表示法”反映事件的现实内容，其余的表示式反映它是$\mathscr{F}_5$中元素，其中右端的表示式简单。

为了便于推广，我们把上述常识改述为：甲(正确地说，ξ)关心的芽事件集是

$$\mathscr{A}_1=\{(\xi=a),(\xi=-a)\} \tag{3.1.10}$$

他关心的全部事件组成事件σ域$\sigma(\mathscr{A}_1)$，通常用$\sigma(\xi)$表示。于是

$$\sigma(\xi)=\sigma(\mathscr{A}_1)=\{\varnothing,(\xi=a),(\xi=-a),\Omega_5\} \tag{3.1.11}$$

今后，称$\sigma(\xi)$为ξ产生的事件σ域，称$(\Omega_5,\sigma(\xi))$为ξ产生的子试验。

(4) ξ的表现概率空间和分布律

称$(\Omega_5,\sigma(\xi),P_5)$为$\xi$的表现概率空间。由定理2.8.1得出概率测度$P_5(\sigma(\xi))$与分布律

$$\begin{pmatrix}\{\xi=-a\} & \{\xi=a\}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}，简记为\begin{pmatrix}-a & a\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix} \tag{3.1.12}$$

相互惟一确定。称(3.1.12)为ξ的分布律。

综合(1)～(4)得出，甲不必关心$(\Omega_5,\mathscr{F}_5,P_5)$的全部统计知识。他只关心子概率空间$(\Omega_5,\sigma(\xi),P_5)$。

同理，乙和丙只关心各自的表现概率空间$(\Omega_5,\sigma(\eta),P_5)$和$(\Omega_5,\sigma(\zeta),P_5)$，其中$\sigma(\eta)$和$\sigma(\zeta)$分别是芽事件集

$$\mathscr{A}_2=\{(\eta=-2a),(\eta=a)\} \tag{3.1.13}$$

$$\mathscr{A}_3=\{(\zeta=-3a),(\zeta=-a),(\zeta=0),(\zeta=2a)\} \tag{3.1.14}$$

生成的事件σ域；$P_5(\sigma(\eta))$与$P_5(\sigma(\zeta))$分别由η与ζ的分布律

$$\begin{pmatrix}-2a & a\\ \frac{1}{3} & \frac{2}{3}\end{pmatrix} \tag{3.1.15}$$

$$\begin{pmatrix} -3a & -a & 0 & 2a \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \tag{3.1.16}$$

确定。同样，丁只关心 s 的表现概率空间$(\Omega_5,\sigma(s),P_5)$，其中 $\sigma(s)=\{\varnothing,\Omega\}$，$P_5(\sigma(s))$由分布律

$$\begin{pmatrix} a \\ 1 \end{pmatrix} \tag{3.1.17}$$

确定。有趣的是，$(\Omega_5,\sigma(s),P_5)$是退化概率空间，它反映博奕中的事实：丁对骰子出现的点数毫无兴趣，他只关心实验 $\mathscr{E}_5$ 是否进行。

3.1.2 分布律的重要性

为了叙述简单和方便，提前把 $\xi,\eta,\cdots$这类量一般化。

规定 1[①] 给定概率空间$(\Omega,\mathscr{F},P)$和实值函数 $X(\omega)$，$\omega\in\Omega$。称 X 为离散型随机变量，如果 X 只取有限或可列个实数值 $x_1,x_2,\cdots$，并且能产生表现概率空间$(\Omega,\sigma(X),P)$。其中

$$\sigma(X)=\sigma\{(X=x_1),(X=x_2),\cdots\} \tag{3.1.18}$$

$P(\sigma(X))$由分布律

$$f:\begin{pmatrix} x_1 & x_2 & \cdots \\ f_1 & f_2 & \cdots \end{pmatrix} \tag{3.1.19}$$

确定。

规定 2 称数值 $\sum\limits_{n\geqslant 1}x_nf_n$ 为 X 的数学期望，记为 $\mathrm{E}X$；称数值 $\sum\limits_{n\geqslant 1}(x_n-\mathrm{E}X)^2f_n$ 为 X 的方差，记为 $\mathrm{D}X$。

应用分布律(3.1.19)可以构造许多的数值，泛称为 X 的数字特征。$\mathrm{E}X$ 和 $\mathrm{D}X$ 是两个最重要的数字特征。特别，当 X 具有特殊

---

① 由于本节不能定义任何概念和符号(参看 2.3 节第(2)段中的脚注)，所以这里采用“规定”一词。事实上，规定只给出概念的不严谨的定义。

分布律

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix} \tag{3.1.20}$$

时数学期望成为

$$\mathrm{E}X = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{3.1.21}$$

它就是常识中的"n 个数 $x_1, x_2, \cdots, x_n$ 的平均值"。这个事实是数学期望获得广泛应用的源头。

分布律的重要性在于：即使$(\Omega, \mathscr{F}, P)$未知，只要分布律已知，就能写出 X 的表现概率空间$(\Omega, \sigma(X), P)$。

事实上，由分布律(3.1.19)中的第一行得出 X 的芽事件集为

$$\mathscr{A} = \{(X = x_1), (X = x_2), \cdots\} \tag{3.1.22}$$

从而 X 产生的事件 σ 域 $\sigma(X)$已知，概率测度 $P(\sigma(X))$由分布律(3.1.19)确定，Ω 是 X 的定义域（如果它无法获得，应用引理 2.4.1可随意地选择一个样本空间）。于是，我们写出了表现概率空间$(\Omega, \sigma(X), P)$。

分布律的重要性还在于：分布律(3.1.19)蕴含 X 的全部统计知识，并产生研究 X 的新工具——数字特征。

容易看出，对于 3.1.1 段中的四人博奕问题，当 a 取不同值时用(3.1.1)规定的 ξ 是不同的，但是这些不同的 ξ 有相同的表现概率空间。因此，ξ 蕴含的统计知识比它的表现概率空间更丰富。丰富部分来自于 ξ 的取值，因为这些值赋予了 $\sigma(\xi)$中事件更多的现实内容（参看(3.1.8)和(3.1.9)）。如何表达这部分统计知识，只能求助于博奕者。

精明的博奕者会提出一系列的问题：输赢规则公平吗？如果不公平，对哪方有利？每局博奕期待赢多少钱？真实的赢钱数与期待值相差多少？等等。

精明的博奕者回答：输赢规则 ξ 和 η 是公平的，因为期待的赢钱数为零；规则 ζ 的期待赢钱数为 $-a$，所以 $a>0$ 时对丙不利，$a<0$ 时对丙有利；规则 s 的讨论类似。其中期待的赢钱数是用(3.1.1)～(3.1.4)算出的平均值(3.1.21)。

概率论比博奕者更精明，它把仅对特殊情形(3.1.20)才有的"平均值"概念推广到一般情形(3.1.19)，产生出数学期望的概念。应用规定 2 与(3.1.12)，(3.1.15)～(3.1.17)容易算出

$$E\xi = E\eta = 0, E\zeta = -a, Es = a \qquad (3.1.23)$$

它和博奕者的经验一致。

至此，概率论找到了表达离散型随机变量统计知识的最简单，最方便的工具——分布律和数字特征。

3.1.3 直观背景：概率空间未知的情形

用 $\mathscr{E}$ 表示实验"天气预报"。气象台每天都发布明天的最高温度为 T，最低温度为 T_1，以及有关雨雪阴晴的天气预报。试问：如何理解天气预报中的这些指标？下面以最高温度 T 为例进行讨论。为叙述准确，不妨认为 T 是明天中午 12 点的气温。

表面上看，T 是一个不确定的量，要事先预言它的值是办不到的。因为气象条件不同时它取不同的数值，而气象条件按偶然的方式变化。鉴于此，人们最初认为这是一类随机会取不同值的变量，简称为随机变量。

自然公理系统为随机变量的研究提供牢固的基础。用 $\mathscr{F}$ 表示实验 $\mathscr{E}$ 产生的全部事件的集合。依据第Ⅱ组公理得出 $\mathscr{F}$ 是事件 σ 域。由引理 2.4.1 得出，存在 $\mathscr{F}$ 的生存随机试验 $(\Omega,\mathscr{F})$。再依据(2.7.4)，能够在 $(\Omega,\mathscr{F})$ 上赋予概率测度 $P(\mathscr{F})$。总之实验 $\mathscr{E}$ 产生概率空间 $(\Omega,\mathscr{F},P)$。

研究的困难在于：$(\Omega,\mathscr{F},P)$ 虽然存在，但无法具体写出它。

克服困难的办法是：放弃获得 $(\Omega,\mathscr{F},P)$ 的全部统计知识的

愿望，转向研究实验 $\mathscr{E}$ 中只与指标 T 有关的统计知识。具体的方法是：

(1) 能够进行"天气预报"的基础是存在概率空间$(\Omega,\mathscr{F},P)$。

虽然$(\Omega,\mathscr{F},P)$不能写出，但它仍然有清楚的直观解释。样本点 ω 代表实验 $\mathscr{E}$ 中的一种气象条件。当 ω 伪出现时，即在这种气象条件下描述天气的所有指标(例如温度，湿度，雨雪阴晴等)都取确定的数值或状态。天气预报的随机性表现为：Ω 中的所有样本点都有机会伪出现，其中有些样本点伪出现的机会大些，有些则小些。使用概念"伪出现"的原因是，$\{\omega\}$往往不是事件；即使是事件，也经常是零概率的。如果希望用"出现"代替"伪出现"，上面的语句应改为：天气预报的随机性表现为 $\mathscr{F}$ 中某些事件(它是一些样本点的集合，即一组气象条件)出现的可能性大些，另一些事件出现的可能性则小些。

(2) T 是定义在$(\Omega,\mathscr{F},P)$上的点函数

有了样本空间 Ω 后能够说，在一种气象条件下指标 T，即最高温度是一个确定的数值。因此 T 是定义在 Ω 上的实值点函数①。

$$T = T(\omega), \omega \in \Omega \tag{3.1.24}$$

(3) T 产生子试验$(\Omega,\sigma(T))$

直观上，"温度 T 大于 a，小于等于 b"，"温度 T 小于等于 x"等都是 $\mathscr{F}$ 中的事件。有趣的是，虽然样本空间 Ω 不能写出，但形成这些事件的样本点的集合却能准确地表示为

"温度 T 大于 a，小于等于 b"

$$=\{\omega \mid a<T(\omega)\leqslant b\} \stackrel{\text{def}}{=\!=} \{a<T\leqslant b\} \tag{3.1.25}$$

$$\text{“温度 } T \text{ 小于等于 } x\text{”}=\{\omega \mid T(\omega)\leqslant x\} \stackrel{\text{def}}{=\!=} \{T\leqslant x\} \tag{3.1.26}$$

---

① 在概率论的发展史中，随机变量的认识由"随机会取不同数值的变量"提高到"它是样本空间上的函数"，这是一个重大的进步。文献[21]在绪论的历史小记中说："样本空间的概念是由米泽斯[12]引进的。这个概念使得有可能把概率的严格数学理论建立在测度论上"。

用 $\sigma(T)$ 表示用事件集合

$$\mathscr{A}=\{(a<T\leqslant b)\mid-\infty<a<b<+\infty\} \qquad (3.1.27)$$

生成的事件 σ 域。常识支持断言：$\sigma(T)$ 包含与 T 有关的全部事件，因此 $\mathscr{A}$ 可以作为 T 的芽事件集①。

今后，称 $\sigma(T)$ 为 T 产生的事件 σ 域，称 $(\Omega,\sigma(T))$ 为 T 产生的子试验。

意义重大的结论是：能够用波莱尔试验 $(\mathbf{R},\mathscr{B})$ 表达 $(\Omega,\sigma(T))$ 的结构。事实上，比较(3.1.27)和(2.5.20)，应用定理 2.4.3得到

$$\sigma(T)=\{(T\in B)\mid B\in\mathscr{B}\} \qquad (3.1.28)$$

其中 $(T\in B)=\{\omega\mid T(\omega)\in B\}$。

(4) T 的表现概率空间和值概率空间

称 $(\Omega,\sigma(T),P)$ 为 T 的表现概率空间。于是，研究 T 的统计知识时不必关心 $(\Omega,\mathscr{F},P)$，只须研究它的子概率空间 $(\Omega,\sigma(T),P)$，其中 $\sigma(T)$ 中事件可以用(3.1.28)一一列举。

应用(3.1.28)，在 $(\mathbf{R},\mathscr{B})$ 上引进集函数

$$P_T(B)=P(T\in B) \quad (B\in\mathscr{B}) \qquad (3.1.29)$$

应用测度论中可测变换的知识易证 $P_T(\mathscr{B})$ 是概率测度[16]。由定理 2.8.2 得出，存在分布函数 $F(x)$，$x\in\mathbf{R}$ 使得 $P_T(\mathscr{B})$ 与 $F_T(x)$ 用关系式

$$F_T(x)=P(T\leqslant x) \quad (x\in\mathbf{R}) \qquad (3.1.30)$$

$$P_T(B)=\int_B \mathrm{d}F_T(x) \quad (B\in\mathscr{B}) \qquad (3.1.31)$$

相互惟一确定。

今后，称柯尔莫哥洛夫概率空间 $(\mathbf{R},\mathscr{B},P_T)$ 为 T 的值概率空

① 参看定理 2.5.7 及随后的议论，以及 2.8 节末注记。

间，且称 $P_T(\mathscr{B})$ 为 T 的概率分布，$F_T(x)$ 为 T 的分布函数。

3.1.4 分布函数的重要性

分布函数的重要性在于：只要已知 T 的分布函数 $F_T(x)$，用(3.1.31)可求出概率分布 $P_T(\mathscr{B})$，从而得到 T 的值概率空间 $(\mathbf{R},\mathscr{B},P_T)$。

值概率空间的重要性在于：① 它通过(3.1.28)和(3.1.29)确定 T 的表现概率空间 $(\Omega,\sigma(T),P)$；② 值概率空间蕴含 T 的全部统计知识。

容易验证，T 和 $kT(k\neq 0,1)$ 有相同的表现概率空间和不同的值概率空间。因此，T 蕴含的统计知识比它的表现概率空间更丰富。丰富部分来自 T 的取值，因为这些值赋予 $\sigma(T)$ 中事件更多的现实内容(参看(3.1.25)与(3.1.26))。

表达丰富部分的统计知识需求助于天气预报的当事者。现实中采用的预报是

$$\text{明天的最高气温为 } \bar{T} \tag{3.1.32}$$

其中 $\bar{T}$ 是一个确定的数值。但是，按照上面的讨论，精确的预报是

$$\text{明天的最高温度是随机变量 } T\text{，它有分布函数 } F_T(x) \tag{3.1.33}$$

事实上，从后一预报中可以得到 T 的全部统计知识。例如，我们能计算出形如(3.1.25)和(3.1.26)中事件的概率。如果假定得到 $P(26.5<T\leqslant 27.5)=0.9$，那么气象台将以 90% 的把握断定最高气温位于区间 $(26.5,27.5]$ 中。

两种预报方式都是合理的(第一种适用于公众；第二种适用于气象研究人员)。常识告诉我们，$\bar{T}$ 是 T 的平均值，它产生的偏差为 $|\bar{T}-T|$。为了准确地理解平均值和偏差，需要把规定 2 拓广为以下规定。

规定 3 设 T 为随机变量，$F_T(x)$ 是它的分布函数，称积分值 $\int_{-\infty}^{+\infty} x\mathrm{d}F_T(x)$ 为 T 的数学期望，记为 $\mathrm{E}T$；称积分值 $\int_{-\infty}^{+\infty}(x-\mathrm{E}T)^2\mathrm{d}F_T(x)$ 为 T 的方差，记为 $\mathrm{D}T$；且称 $\sqrt{\mathrm{D}T}$ 为 T 的标准差或均方差。

于是，第一种预报方法中的 $\overline{T}$ 是 T 的数学期望，产生的偏差 $|T-\overline{T}|$ 与 $\sqrt{\mathrm{D}T}$ 紧密相关。随着人们科学素质的提高，可以设想出现第三种方式的预报：

$$\text{明天的最高气温为 } \overline{T}\text{，标准差为 } \varepsilon \tag{3.1.34}$$

3.1.5 随机变量在应用中的普遍性

用 X 表示天气预报中涉及“雨雪阴晴”的指标。为简单起见，不妨假定 X 只取四种状态：“雨”、“雪”、“阴”、“晴”(即不再区分大雨，小雨等状态)。类似于获得(3.1.24)的讨论得出 X 是定义在 Ω 上以状态为值的函数 $X(\omega)$，$\omega\in\Omega$。如果我们把状态数值化，例如采用数值化表格

状　　态	雨	雪	阴	晴
对应数值	1	2	3	4

(3.1.35)

后 X 成为定义在$(\Omega,\mathscr{F},P)$上的离散型随机变量，它的分布律为

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ f_1 & f_2 & f_3 & f_4 \end{pmatrix} \tag{3.1.36}$$

于是，预报“雨雪阴晴”的问题归结为寻找分布律(3.1.36)。如果已知(3.1.36)，气象台会发布“明天有雨的概率是 f_1”的天气预报。

顺便指出，如果人为地认为 X 的值域为 $\mathbf{R}$，即认为 X 可以取所有的实数值，并规定事件

$$\{\omega \mid X(\omega)\neq 1,2,3,4\}$$

(如果是规定 1 中的 X,则为$\{\omega|X(\omega)\neq x_n, n=1,2,3,\cdots\}$)的概率为零,那么离散型随机变量成为 3.1.3 中随机变量的特殊情形。

3.1.6 随机变量在理论中的普遍性

随机变量 $X(\omega)$产生表现概率空间$(\Omega,\sigma(X),P)$。反问题是:设$(\Omega,\mathscr{F}_0,P)$是$(\Omega,\mathscr{F},P)$的子概率空间,试问是否存在$(\Omega,\mathscr{F},P)$上的随机变量 $X(\omega)$,使得 $X(\omega)$的表现概率空间是$(\Omega,\mathscr{F}_0,P)$?

可以证明,当 $\mathscr{F}_0$ 是可分的事件 σ 域时,问题有肯定的答案[24.16]。

3.2 随机变量的基本概念

定义 3.2.1 给定概率空间$(\Omega,\mathscr{F},P)$和实值函数 $X(\omega)$,$\omega\in\Omega$。称 X 是$(\Omega,\mathscr{F},P)$上的随机变量,如果对任意的实数 x 成立

$$(X\leqslant x)\overset{\text{def}}{=\!=}\{\omega|X(\omega)\leqslant x\}\in\mathscr{F}\tag{3.2.1}$$

定义 3.2.2 给定概率空间$(\Omega,\mathscr{F},P)$和波莱尔试验$(\mathbf{R},\mathscr{B})$。设 $X(\omega)$是定义在 Ω 上取值于 $\mathbf{R}$ 的函数。如果对任意的 $B\in\mathscr{B}$ 成立

$$(X\in B)\overset{\text{def}}{=\!=}\{\omega|X(\omega)\in B\}\in\mathscr{F}\tag{3.2.2}$$

则称 X 是$(\Omega,\mathscr{F},P)$上的随机变量。

众所周知,两个定义等价。给出两个定义的目的是希望从两个角度理解随机变量。用 $\mathscr{A}$ 和 $\sigma(X)$分别表示(3.2.1)和(3.2.2)中所有事件组成的集合,即

$$\mathscr{A}=\{(X\leqslant x)|x\in\mathbf{R}\}\tag{3.2.3}$$

$$\sigma(X)=\{(X\in B)|B\in\mathscr{B}\}\tag{3.2.4}$$

那么$\sigma(X)$是$\mathscr{F}$的子事件σ域，$\mathscr{A}$是生成$\sigma(X)$的芽事件集。依据3.1节显示的直观背景和2.8节末注记，我们引进以下定义。

定义3.2.3 称$\sigma(X)$为X产生的事件σ域，它是涉及X的所有事件组成的集合。

随机变量的研究包含两方面的内容：(1) 在$(\Omega,\mathscr{F},P)$已知或未知的情况下寻找X蕴含的统计知识和统计规律；(2) 研究随机变量之间的关系，特别是关系中的统计知识和统计规律。

本节讨论第(1)项内容，后两节讨论第(2)项内容。

3.2.1 表现概率空间和数字特征

定义3.2.4 称$(\Omega,\sigma(X),P)$为X的表现概率空间。称$(\Omega,\sigma(X))$为X产生的随机试验。

X的统计知识包括表现概率空间，并且比它更丰富，因为X的取值赋予$\sigma(X)$中事件更多的现实内容。于是人们应用X的取值构造许多的新数值，并用这些新数值反映丰富部分的统计知识。这样的新数值被称为X的数字特征。这里只介绍两个最重要的数字特征。

定义3.2.5 设X是$(\Omega,\mathscr{F},P)$上的随机变量。如果积分$\int_\Omega |X(\omega)|P(\mathrm{d}\omega) < +\infty$，则称$\int_\Omega X(\omega)P(\mathrm{d}\omega)$为$X$的数学期望(或期望，或均值)，记为$\mathrm{E}X$。即

$$\mathrm{E}X = \int_\Omega X(\omega)P(\mathrm{d}\omega) \tag{3.2.5}$$

定义3.2.6 如果随机变量$(X-\mathrm{E}X)^2$存在数学期望，则称期望值为X的方差，记为$\mathrm{D}X$。即

$$\mathrm{D}X=\mathrm{E}(X-\mathrm{E}X)^2=\int_\Omega [X(\omega)-\mathrm{E}X]^2P(\mathrm{d}\omega) \tag{3.2.6}$$

并称$\sqrt{\mathrm{D}X}$为X的标准差，或均方差。

对任意的$A\in\mathscr{F}$，$\chi_A(\omega)$是A的示性函数，通过简单的积分计

算得到

$$\mathrm{E}\chi_A = P(A), \mathrm{D}\chi_A = P(A)P(\bar{A}) \tag{3.2.7}$$

于是，事件的概率等于它的示性函数的数学期望。

用期望和方差可以反映随机变量的许多统计规律。这方面的一个经典结果为以下定理。

定理 3.2.1（切比雪夫(P. L. Chebyshev)不等式）[①] 设随机变量 X 存在数学期望和方差。那么对任意的 $\varepsilon>0$ 成立

$$P(|X - \mathrm{E}X| \geqslant \varepsilon) \leqslant \frac{\mathrm{D}X}{\varepsilon^2} \tag{3.2.8}$$

特别，取 $\varepsilon=10\sqrt{\mathrm{D}X}$，由上式得出

$$P(\mathrm{E}X-10\sqrt{\mathrm{D}X}<X<\mathrm{E}X+10\sqrt{\mathrm{D}X})\geqslant 0.99 \tag{3.2.9}$$

于是得到统计规律：对任意的随机变量 X，能够以 99%的把握断定 X 将出现在区间$(\mathrm{E}X-10\sqrt{\mathrm{D}X}, \mathrm{E}X+10\sqrt{\mathrm{D}X})$之中。

3.2.2 值概率空间和分布函数

设$(\mathbf{R},\mathscr{B})$是定义 3.2.2 中的波莱尔试验。应用(3.2.4)引进

$$P_X(B) = P(X \in B) \quad (B \in \mathscr{B}) \tag{3.2.10}$$

易证 $P_X(\mathscr{B})$是概率测度。由定理 2.8.2 得出，存在分布函数 $F_X(x)$，$x\in\mathbf{R}$使得 $P_X(\mathscr{B})$与 $F_X(x)$用关系式

$$F_X(x) = P_X((-\infty,x]) = P(X \leqslant x)(x \in \mathbf{R}) \tag{3.2.11}$$

$$P_X(B) = \int_B \mathrm{d}F_X(x) \qquad (B \in \mathscr{B}) \tag{3.2.12}$$

相互惟一确定。

定义 3.2.7 称$(\mathbf{R},\mathscr{B},P_X)$为 X 的值概率空间，且称 $P_X(\mathscr{B})$为 X 的概率分布，$F_X(x)$为 X 的分布函数。

---

① 本章是简介，给出的定理皆不证明。

测度论中可测变换理论保证 X 在$(\Omega,\mathscr{F},P)$上的性质和表达式与$(\mathbf{R},\mathscr{B},P_X)$的对应性质和表达式可以互相转换。即是说，值概率空间包含随机变量的全部统计知识。特别，(3.2.5)和(3.2.6)在值概率空间中分别成为

$$\mathrm{E}X=\int_{\mathbf{R}}xP_X(\mathrm{d}x)=\int_{\mathbf{R}}x\mathrm{d}F_X(x) \tag{3.2.13}$$

$$\mathrm{D}X=\int_{\mathbf{R}}(x-\mathrm{E}X)^2P_X(\mathrm{d}x)=\int_{\mathbf{R}}(x-\mathrm{E}X)^2\mathrm{d}F_X(x) \tag{3.2.14}$$

其中右端是 L-S 积分。

于是，概率论找到表达随机变量全部统计知识的最方便，最简单的工具——分布函数及由它产生的数字特征。

分布函数和值概率空间在随机变量的研究中具有重大的意义。原因是：

(1) 研究随机变量时不必拘泥于写出它所在的概率空间$(\Omega,\mathscr{F},P)$。这个事实具有巨大的实用价值，因为应用中的随机变量往往不能写出它所在的概率空间$(\Omega,\mathscr{F},P)$。

事实上，假定 X 的分布函数 $F_X(x)$已知，则值概率空间$(\mathbf{R},\mathscr{B},P_X)$已知。那么，用(3.2.4)可以完整地表达 $\sigma(X)$，用(3.2.10)可以得到 $P(\sigma(X))$，从而在引理 2.4.1 的意义下得到 X 的表现概率空间$(\Omega,\sigma(X),P(\sigma(X)))$。于是，可以用$(\Omega,\sigma(X),P(\sigma(X)))$代替$(\Omega,\mathscr{F},P)$，并且不会丧失 X 的任何统计知识。

(2) 它把概率论的公式由 3.2.1 段中的测度论形式转变为实变函数的形式，给概率计算带来许多方便。例如，我们经常用(3.2.12)～(3.2.14)计算有关事件的概率、X 的期望和方差。

(3) 在相当广泛的条件下 L-S 积分是级数与黎曼积分。如果限定随机变量是离散型的或连续型的(它们能满足大部分应用的要求)，本段中的公式成为级数形式或黎曼积分形式的公式。于是，

我们只须初等微积分就可以在值概率空间中研究随机变量 X，获得 X 的许多性质。通常，这些性质是用有关分布函数或数字特征的公式来表达，这些性质的概率意义则在表现概率空间中解释。事实上，许多概率论的入门书籍就是这样处理的。

3.2.3 注记：概率分析方法

为了更好地掌握上述方法的实质，以便把它们推广到随机变量族，让我们泛泛地谈论研究方法的问题。

3.2.1 段是在概率空间 $(\Omega,\mathscr{F},P)$ 中对随机变量 $X(\omega)$ 进行直接的研究；3.2.2 段是在值概率空间 $(\mathbf{R},\mathscr{B},P_X)$ 中以分布函数 $F_X(x)$ 为工具对 $X(\omega)$ 进行间接的研究。因此，前者称为概率方法，后者称为分析方法。

既然是为了推广，姑且承认凡是关于 X 的议论也适用于随机变量族 $X_T \overset{\text{def}}{=} \{X_t(\omega)\mid t\in T\}$（事实上，它不是完全正确的结论）。并且为了直观意义明显，假定指标集 T 是时间 $(-\infty,+\infty)$。

（1）概率方法

本方法是在 $(\Omega,\mathscr{F},P)$ 上对 X_T 进行直接的研究。由于 X_T 是现实中存在的量，有清晰的直观背景，因而用 X_T 产生的概念、表达式和数学演算也有明显的直观形象，容易理解。例如，

① 表现概率空间 $(\Omega,\sigma(X_T),P)$ 表达出与 X_T 有关的全部事件及其概率。

② 式(3.2.11)、(3.2.5)和(3.2.6)表示 X_T 中单个随机变量的分布函数、期望和方差。把它们作为整体后得到 X_T 的分布函数族，定义在 T 上的期望函数和方差函数。

③ 依据直观背景，人们关心 X_T 首次到达区域 G 的时刻 T_G，即

$$T_G(\omega)=\inf\{t\mid X_t(\omega)\in G\} \qquad (3.2.15)$$

称 T_G 为 G 的首达时。同理，称

$$\varepsilon_G(\omega) = \sup\{t \mid X_t(\omega) \in G\} \qquad (3.2.16)$$

为 G 的末离时，它是 X_T 最后离开 G 的时刻。

显然，理论和应用都要求人们研究 T_G 和 ε_G 的统计知识。

④ 不妨把 X_T 确定的随机变量(例如 T_G 和 ε_G)泛称为随机泛函，试寻找有实用价值的随机泛函，并研究它们的统计知识。

⑤ 探讨 X_t 与时间 t 的依赖关系。特别，它关于 t 是否连续、可微、可积？这里的连续、可微和可积又如何理解？

概率方法的优点是直观性和形象性。它促使人们不断地提出新问题，新任务，寻找新方法。因而使概率论充满了生机和活力，获得广泛的应用。

概率方法的特点是，为了把具有直观形象的性质及其演变理论化，需要用抽象的测度论方法把性质转化为公式和把演变转化为对 X_T 的演算。因此熟悉测度论是必要的。

(2) 分析方法

分析方法的实质是，首先依据直观背景把概率问题转化为涉及分布函数的分析问题；然后用分析方法求出问题的答案；最后对分析形式的答案进行概率解释。

分析方法的优点是运用成熟的分析工具研究概率论。特别，在初等微积分基础上就能够深入地研究具有广泛应用的概率论内容。

分布函数不是现实中出现的量。它是人们为了表达随机变量统计知识时引进的工具。因此，分析方法中前后两个转化要求人们了解和掌握概率论的直观背景。历史上，在样本空间和测度论出现之前，随机变量是一个含糊的、难于精确定义的一种量。当时的科学家只能凭据敏锐的概率论直觉，应用分布函数对它们进行研究，并且获得大量的成果，形成现今所称的“分析概率论”。

现代概率论用概率分析方法研究随机变量。即是说，既用分析方法，又用概率方法，两者相辅相成地进行研究，极大地发挥出两

种方法的效力，使概率论得到蓬勃的发展。

分析数学也需要不断的发展。两种方法的揉合不仅使我们能够用分析方法研究概率论，而且能够用概率方法研究分析理论。目前已经出现一些用概率方法发现和证明分析学中新成果的工作，它们为分析数学开拓出一片新天地。

3.3 随机向量的基本概念

定义 3.3.1 给定概率空间$(\Omega,\mathscr{F},P)$和向量值函数$\boldsymbol{X}(\omega)\overset{\text{def}}{=}(X_1(\omega),X_2(\omega),\cdots,X_n(\omega))$。称$\boldsymbol{X}$是$(\Omega,\mathscr{F},P)$上的$n$维随机向量，如果它的每个分量都是随机变量。即对任意的实数x_i有

$$(X_i\leqslant x_i)\overset{\text{def}}{=}\{\omega\,|\,X_i(\omega)\leqslant x_i\}\in\mathscr{F}\quad(i=1,2,\cdots,n)\tag{3.3.1}$$

容易验证，条件(3.3.1)等价于用柱事件表达的条件：对任意的$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$有

$$\begin{aligned}(\boldsymbol{X}\leqslant\boldsymbol{x})&\overset{\text{def}}{=}(X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n)\\&=\{\omega\,|\,X_1(\omega)\leqslant x_1,X_2(\omega)\leqslant x_2,\cdots,X_n(\omega)\leqslant x_n\}\in\mathscr{F}\end{aligned}\tag{3.3.2}$$

定义 3.3.2 给定概率空间$(\Omega,\mathscr{F},P)$和n维波莱尔试验$(\mathbf{R}^n,\mathscr{B}^n)$。设$\boldsymbol{X}(\omega)=(X_1(\omega),X_2(\omega),\cdots,X_n(\omega))$是定义在$\Omega$上取值于$\mathbf{R}^n$的函数，如果对任意的$B\in\mathscr{B}^n$有

$$(\boldsymbol{X}\in B)\overset{\text{def}}{=}\{\omega\,|\,\boldsymbol{X}(\omega)\in B\}\in\mathscr{F}\tag{3.3.3}$$

则称$\boldsymbol{X}$是$(\Omega,\mathscr{F},P)$上的n维随机向量。

众所周知，两个定义等价。用$\mathscr{A}_n$和$\sigma(\boldsymbol{X})$分别表示(3.3.2)和(3.3.3)中所有事件的集合，即

$$\mathscr{A}_n=\{(\boldsymbol{X}\leqslant\boldsymbol{x})\,|\,\boldsymbol{x}\in\mathbf{R}^n\}\tag{3.3.4}$$

$$\sigma(\boldsymbol{X})=\{(\boldsymbol{X}\in B)\,|\,B\in\mathscr{B}^n\}\tag{3.3.5}$$

那么$\sigma(\boldsymbol{X})$是$\mathscr{F}$的子事件σ域，$\mathscr{A}_n$是生成$\sigma(\boldsymbol{X})$的芽事件集。类似于随机变量的情形，我们引进

定义 3.3.3 称$\sigma(\boldsymbol{X})$为$\boldsymbol{X}$产生的事件σ域，它是涉及$\boldsymbol{X}$的所有事件组成的集合。

现在能够解释为什么要引进等价的两个定义。比较定义3.3.2和3.2.2即可猜测：把$\boldsymbol{X}$视为整体时随机变量的研究方法和结论大致可以推广到随机向量。这个想法引出第(1)、(2)段的内容。比较定义3.3.1和3.2.1看出，用这个定义研究n个随机变量X_1，X_2，…，X_n之间的关系是方便的。这个想法引出第(3)～(7)段的内容。

3.3.1 表现概率空间和数字特征

定义 3.3.4 称$(\Omega,\sigma(\boldsymbol{X}),P)$为$\boldsymbol{X}$的表现概率空间。称$(\Omega,\sigma(\boldsymbol{X}))$为$\boldsymbol{X}$产生的随机试验。

定义 3.3.5 设$\boldsymbol{X}$是$(\Omega,\mathscr{F},P)$上的随机向量。如果积分$\int_\Omega \boldsymbol{X}(\omega)P(\mathrm{d}\omega)$存在①，则称积分值为$\boldsymbol{X}$的数学期望(或期望向量，或期望，或均值)，记为$\mathrm{E}\boldsymbol{X}$，即

$$\mathrm{E}\boldsymbol{X}=\int_\Omega \boldsymbol{X}(\omega)P(\mathrm{d}\omega) \tag{3.3.6}$$

定义 3.3.6 如果积分$\int_\Omega[\boldsymbol{X}(\omega)-\mathrm{E}\boldsymbol{X}]^{\mathrm{T}}[\boldsymbol{X}(\omega)-\mathrm{E}\boldsymbol{X}]P(\mathrm{d}\omega)$存在②，则称积分值为$\boldsymbol{X}$的协方差矩阵，记为$\mathrm{cov}(\boldsymbol{X},\boldsymbol{X})$。即

$$\mathrm{cov}(\boldsymbol{X},\boldsymbol{X})=\int_\Omega[\boldsymbol{X}(\omega)-\mathrm{E}\boldsymbol{X}]^{\mathrm{T}}[\boldsymbol{X}(\omega)-\mathrm{E}\boldsymbol{X}]P(\mathrm{d}\omega) \tag{3.3.7}$$

---

① 向量的积分$\int_\Omega \boldsymbol{X}(\omega)P(\mathrm{d}\omega)$被规定为$\left(\int_{\Omega_1}X_1(\omega)P(\mathrm{d}\omega),\int_{\Omega_2}X_2(\omega)P(\mathrm{d}\omega),\cdots,\int_\Omega X_n(\omega)P(\mathrm{d}\omega)\right)$。积分存在是指它的分量都是绝对可积的。

② 矩阵的积分被规定为对它的每个分量进行积分。积分存在是指每个分量皆绝对可积。

由向量值函数的积分得知，$\mathrm{E}\boldsymbol{X}$ 是一个数值向量，即

$$\mathrm{E}\boldsymbol{X} = (\mathrm{E}X_1, \mathrm{E}X_2, \cdots, \mathrm{E}X_n) \tag{3.3.8}$$

由矩阵值函数的积分得知，$\mathrm{cov}(\boldsymbol{X},\boldsymbol{X})$是一个 $n\times n$ 阶数值矩阵。它是

$$\mathrm{cov}(\boldsymbol{X},\boldsymbol{X}) = \mathrm{E}(\boldsymbol{X} - \mathrm{E}\boldsymbol{X})^{\mathrm{T}}(\boldsymbol{X} - \mathrm{E}\boldsymbol{X})$$

$$= \begin{pmatrix} \mathrm{cov}(X_1,X_1) & \mathrm{cov}(X_1,X_2) & \cdots & \mathrm{cov}(X_1,X_n) \\ \mathrm{cov}(X_2,X_1) & \mathrm{cov}(X_2,X_2) & \cdots & \mathrm{cov}(X_2,X_n) \\ \vdots & \vdots & & \vdots \\ \mathrm{cov}(X_n,X_1) & \mathrm{cov}(X_n,X_2) & \cdots & \mathrm{cov}(X_n,X_n) \end{pmatrix} \tag{3.3.9}$$

其中

$$\mathrm{cov}(X_i,X_j) = \mathrm{E}(X_i - \mathrm{E}X_i)(X_j - \mathrm{E}X_j)$$

$$= \int_\Omega [X_i(\omega) - \mathrm{E}X_i][X_j(\omega) - \mathrm{E}X_j]P(\mathrm{d}\omega) \tag{3.3.10}$$

当 $i\neq j$ 时称 $\mathrm{cov}(X_i,X_j)$为 X_i 与 X_j 的协方差；当 $i=j$ 时 $\mathrm{cov}(X_i, X_i)=\mathrm{D}X_i$。

3.3.2 值概率空间和分布函数

设$(\mathbf{R}^n,\mathscr{B}^n)$是定义 3.3.2 中的 n 维波莱尔试验。应用(3.3.5)可以在$(\mathbf{R}^n,\mathscr{B}^n)$上引进集函数

$$P_{\boldsymbol{X}}(B) = P(\boldsymbol{X} \in B), B \in \mathscr{B}^n \tag{3.3.11}$$

易证 $P_{\boldsymbol{X}}(\mathscr{B}^n)$是概率测度。由定理 2.8.3 得出，存在 n 元分布函数 $F_{\boldsymbol{X}}(\boldsymbol{x}),\boldsymbol{x}\in\mathbf{R}^n$ 使得 $P_{\boldsymbol{X}}$ 与 $F_{\boldsymbol{X}}$ 用关系式

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = P_{\boldsymbol{X}}\left(\prod_{i=1}^{n}(-\infty,x_i]\right) = P(\boldsymbol{X} \leqslant \boldsymbol{x}) \quad (\boldsymbol{x} \in \mathbf{R}^n) \tag{3.3.12}$$

$$P_{\boldsymbol{X}}(B) = \int_B \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \quad (B \in \mathscr{B}^n) \tag{3.3.13}$$

相互惟一确定。其中 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$。

定义 3.3.7 称$(\mathbf{R}^n,\mathscr{B}^n,P_X)$为 $\boldsymbol{X}$ 的值概率空间，且称 $P_X(\mathscr{B}^n)$为 $\boldsymbol{X}$ 的概率分布，$F_X(\boldsymbol{x})$为 $\boldsymbol{X}$ 的分布函数，或 $X_1,X_2,\cdots,X_n$ 的 n 维联合分布函数。

测度论中可测变换的理论保证：$\boldsymbol{X}$ 在$(\Omega,\mathscr{F},P)$上的性质和表达式与$(\mathbf{R}^n,\mathscr{B}^n,P_X)$的对应性质和表达式之间可以互相转换。即是说，值概率空间包含 $\boldsymbol{X}$ 的全部统计知识。特别，期望向量与协方差矩阵在值概率空间中的表达式分别是

$$\mathrm{E}\boldsymbol{X}=\int_{\mathbf{R}^n}\boldsymbol{x}P_X(\mathrm{d}\boldsymbol{x})=\int_{\mathbf{R}^n}\boldsymbol{x}\mathrm{d}F_X(\boldsymbol{x}),\tag{3.3.14}$$

$$\begin{aligned}\operatorname{cov}(\boldsymbol{X},\boldsymbol{X})&=\int_{\mathbf{R}^n}(\boldsymbol{x}-\mathrm{E}\boldsymbol{X})^{\mathrm{T}}(\boldsymbol{x}-\mathrm{E}\boldsymbol{X})P_X(\mathrm{d}\boldsymbol{x})\\&=\int_{\mathbf{R}^n}(\boldsymbol{x}-\mathrm{E}\boldsymbol{X})^{\mathrm{T}}(\boldsymbol{x}-\mathrm{E}\boldsymbol{X})\mathrm{d}F_X(\boldsymbol{x}),\end{aligned}\tag{3.3.15}$$

其中右端是 n 维 L-S 积分。

于是，随机向量找到了表达自己全部统计知识最方便、最简单的工具——分布函数和它产生的数字特征。

分布函数和值概率空间在随机向量的研究中具有重大的意义。上节对随机变量陈述的理由也适用于随机向量。

3.3.3 子随机向量和边缘分布函数

给定随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$。任取正整数 $1\leqslant i_1<i_2<\cdots<i_r\leqslant n$，$1\leqslant r\leqslant n$。称 $\xi\overset{\text{def}}{=}(X_{i_1},X_{i_2},\cdots,X_{i_r})$为 $\boldsymbol{X}$ 的 r 维子随机向量，简称为子向量。

定义 3.3.8 **称 ξ 的值概率空间$(\mathbf{R}^r,\mathscr{B}^r,P_\xi)$为 $\boldsymbol{X}$ 的 r 维边缘值概率空间；称 ξ 的概率分布 $P_\xi(\mathscr{B}^r)$为 $\boldsymbol{X}$ 的 r 维边缘概率分布；称 ξ 的分布函数 $F_\xi(x_{i_1},x_{i_2},\cdots,x_{i_r})$为 $\boldsymbol{X}$ 的 r 维边缘分布函数。**

为了不使记号繁杂，不失一般性可以用 $\xi=(X_1,X_2,\cdots,X_r)$作

为子向量的代表。对任意的 $B\in\mathscr{B}^r$,构造集合

$$B^* = B\times\prod_{i=r+1}^{n}\mathbf{R} = \{(x_1,x_2,\cdots,x_n)\mid(x_1,x_2,\cdots,x_r),\\ \in B, x_{r+1},\cdots,x_n\in\mathbf{R}\} \tag{3.3.16}$$

显然 $B^*\in\mathscr{B}^n$,并且$(\xi\in B)$和$(\boldsymbol{X}\in B^*)$表示 $\sigma(\xi)$中同一个事件。因此

$$P_\xi(B) = P(\xi\in B) = P(\boldsymbol{X}\in B^*) = P_{\boldsymbol{X}}(B^*) \tag{3.3.17}$$

特别,取 $B=\prod_{i=1}^{r}(-\infty,x_i]$时它成为:

定理 3.3.1 $\boldsymbol{X}$ 的边缘分布函数由 $\boldsymbol{X}$ 的联合分布函数 $F_{\boldsymbol{X}}(\boldsymbol{x})$ 惟一确定。特别,

$$F_\xi(x_1,x_2,\cdots,x_r) = \lim_{x_{r+1},\cdots,x_n\to+\infty} F_{\boldsymbol{X}}(x_1,x_2,\cdots,x_n) \tag{3.3.18}$$

应用边缘分布函数可以简化涉及 $F_{\boldsymbol{X}}$ 的许多表达式。例如,(3.3.14)和(3.3.15)的分量的表达式分别为

$$\mathrm{E}X_i = \int_{\mathbf{R}^n} x_i \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \tag{3.3.19}$$

$$\mathrm{cov}(X_i,X_j) = \int_{\mathbf{R}^n}(x_i-\mathrm{E}X_i)(x_j-\mathrm{E}X_j)\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \tag{3.3.20}$$

现在可以分别简化为

$$\mathrm{E}X_i = \int_{\mathbf{R}} x_i \mathrm{d}F_{X_i}(x_i) \tag{3.3.21}$$

$$\mathrm{cov}(X_i,X_j) = \int_{\mathbf{R}^2}(x_i-\mathrm{E}X_i)(x_j-\mathrm{E}X_j)\mathrm{d}F_{(X_i,X_j)}(x_i,x_j) \tag{3.3.22}$$

3.3.4 独立性

定义 3.3.9 设 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_m$ 是$(\Omega,\mathscr{F},P)$上的 m 个随机向量。称 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_m$ 是相互独立的,如果它们产生的子随机试验$(\Omega,\sigma(\boldsymbol{X}_i))(i=1,2,\cdots,m)$是相互独立的。

应用定理 2.9.3 和 3.3.1 立即推出

定理 3.3.2 随机向量 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_m$ 相互独立的充分必要条件是

$$F_{(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_m)}(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m)=\prod_{i=1}^{m}F_{\boldsymbol{X}_i}(\boldsymbol{x}_i) \qquad (3.3.23)$$

特别,当 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_m$ 皆退化为随机变量时得到 m 个随机变量相互独立的概念。

3.3.5 各种形式的条件概率

第Ⅴ组公理在$(\Omega,\mathscr{F},P)$上引进四种条件概率

$$P(\mathscr{F}_1|\mathscr{F}_2),P(C|\mathscr{F}_2),P(\mathscr{F}_1|D),P(C|D) \qquad (3.3.24)$$

其中 $C,D\in\mathscr{F}$,$\mathscr{F}_1$ 和 $\mathscr{F}_2$ 是 $\mathscr{F}$ 的子事件 σ 域。

设 $\boldsymbol{X}$ 和 $\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_m)$是$(\Omega,\mathscr{F},P)$上的随机向量。令 $\mathscr{F}_1=\sigma(\boldsymbol{X})$,$\mathscr{F}_2=\sigma(\boldsymbol{Y})$,现在对条件概率

$$P(\sigma(\boldsymbol{X})|\sigma(\boldsymbol{Y})),P(C|\sigma(\boldsymbol{Y})),P(\sigma(\boldsymbol{X})|D) \qquad (3.3.25)$$

产生特殊的兴趣。为了书写简明,我们把它们分别简写为

$$P(\boldsymbol{X}|\boldsymbol{Y}),P(C|\boldsymbol{Y}),P(\boldsymbol{X}|D) \qquad (3.3.26)$$

并分别简称为 $\boldsymbol{X}$ 关于 $\boldsymbol{Y}$ 的条件概率(测度族),事件 C 关于 $\boldsymbol{Y}$ 的条件概率(值族),$\boldsymbol{X}$ 关于事件 D 的条件概率(测度)。

(1) $\boldsymbol{X}$ 关于 D 的条件概率分布和条件分布函数

在概率空间$(\Omega,\mathscr{F},P)$中,$P(\boldsymbol{X}|D)$是子试验$(\Omega,\sigma(\boldsymbol{X}))$上的一个概率测度

$$P(C|D),C\in\sigma(\boldsymbol{X}) \qquad (3.3.27)$$

转向 $\boldsymbol{X}$ 的值概率空间$(\mathbf{R}^n,\mathscr{B}^n,P_X)$。应用(3.3.27)可以在$(\mathbf{R}^n,\mathscr{B}^n)$上引进一个新概率测度 $P_X(\mathscr{B}^n|D)$,

$$P_X(A|D)\overset{\text{def}}{=\!=}P(\boldsymbol{X}\in A|D),A\in\mathscr{B}^n \qquad (3.3.28)$$

对任意的$(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$,令

$$F_X(x_1,x_2,\cdots,x_n|D) \stackrel{\text{def}}{=} P_X\left(\prod_{i=1}^{n}(-\infty,x_i]|D\right) \quad (3.3.29)$$

由定理 2.8.3 知，它是 $P_X(A|D), A\in\mathscr{B}^n$ 的分布函数。

定义 3.3.10 称 $P_X(A|D), A\in\mathscr{B}^n$ 为 X 关于 D 的条件概率分布；称 $F_X(x_1,x_2,\cdots,x_n|D),(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$ 为 X 关于 D 的条件分布函数。

(2) 事件 C 关于 Y 的条件密度

在概率空间$(\Omega,\mathscr{F},P)$中，$P(C|Y)$是 C 关于 Y 的条件概率。它是定义在子试验$(\Omega,\sigma(Y))$上的集函数

$$P(C|D), D\in\sigma(Y) \quad (3.3.30)$$

由定理 2.10.1 得出，它存在条件密度 $p(C|\sigma(Y))(\omega)$。现在把它简记为

$$p(C|Y)(\omega), \omega\in\Omega \quad (3.3.31)$$

转向 Y 的值概率空间$(\mathbf{R}^m,\mathscr{B}^m,P_Y)$。由于(3.3.31)的函数关于 ω 是 $\sigma(Y)$可测的，由可测变换的知识得出，存在$(\mathbf{R}^m,\mathscr{B}^m,P_Y)$上的波莱尔函数 $\varphi(y), y\in\mathbf{R}^m$ 使得

$$p(C|Y)(\omega)=\varphi(Y(\omega)) \quad (P_Y\text{-a.s.}) \quad (3.3.32)$$

并且在不计 P_Y 零概率集的意义下 φ 是惟一的。

依照概率论的习惯，$\varphi(y)$用专有记号 $p(C|Y)(y), y\in\mathbf{R}^m$ 代替[①]。注意到(3.3.31)后引进：

定义 3.3.11 称 $p(C|Y)$为事件 C 关于 Y 的条件密度。如果它的自变量为 y，则称之为柯尔莫哥洛夫形式的条件密度；如果它的自变量为 ω，则称之为杜布(J. L. Doob)形式的条件密度。

(3) X 关于子试验$(\Omega,\mathscr{F}_2)$的条件密度—概率，条件密度—概率分布和条件密度—分布函数

① 在这个记号下(3.3.32)成为一个有点别扭的等式 $p(C|Y)(\omega)=p(C|Y)(Y(\omega))$。因此有的书籍中采用的专有记号是 $p(C|Y=y)$。

在概率空间$(\Omega,\mathscr{F},P)$中条件概率$P(\sigma(\boldsymbol{X})|\mathscr{F}_2)$可以简写为$P(\boldsymbol{X}|\mathscr{F}_2)$，简称为$\boldsymbol{X}$关于$\mathscr{F}_2$的条件概率(测度族)，它是二元集函数

$$P(C|D),C\in\sigma(\boldsymbol{X}),D\in\mathscr{F}_2^* \qquad (3.3.33)$$

应用定理2.10.4可证，条件概率$P(\boldsymbol{X}|\mathscr{F}_2)$存在条件密度—概率

$$p(C|\omega;\mathscr{F}_2),(C,\omega)\in\sigma(\boldsymbol{X})\times\Omega \qquad (3.3.34)$$

转向$\boldsymbol{X}$的值概率空间$(\mathbf{R}^n,\mathscr{B}^n,P_X)$。应用(3.3.5)和(3.3.34)可以定义新的集一点函数

$$p_X(A|\omega;\mathscr{F}_2)\xlongequal{\text{def}}p(\boldsymbol{X}\in A|\omega;\mathscr{F}_2),(A,\omega)\in\mathscr{B}^n\times\Omega \qquad (3.3.35)$$

容易验证它具有下列3个性质：

① 对固定的A，p_X关于ω是$\mathscr{F}_2$可测函数；

② 对固定的ω，p_X关于A是$(\mathbf{R}^n,\mathscr{B}^n)$上的概率测度；

③ 对任意的$A\in\mathscr{B}^n$，$D\in\mathscr{F}_2^*$成立

$$P(\boldsymbol{X}\in A|D)=P_X(A|D)=\frac{1}{P(D)}\int_D p_X(A|\omega;\mathscr{F}_2)P(\mathrm{d}\omega) \qquad (3.3.36)$$

对任意的$(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$，$\omega\in\Omega$，令

$$F_{X|\mathscr{F}_2}(x_1,x_2,\cdots,x_n|\omega)\xlongequal{\text{def}}p_X\left(\prod_{i=1}^n(-\infty,x_i]|\omega;\mathscr{F}_2\right) \qquad (3.3.37)$$

显然，对每个固定的ω，它是$p_X(A|\omega;\mathscr{F}_2)$的分布函数。

定义3.3.12 ① 称$p(C|\omega;\mathscr{F}_2)$，$(C,\omega)\in\sigma(\boldsymbol{X})\times\Omega$为$\boldsymbol{X}$关于$(\Omega,\mathscr{F}_2)$(或$\mathscr{F}_2$)的条件密度—概率。

② 称$p_X(A|\omega;\mathscr{F}_2)$，$(A,\omega)\in\mathscr{B}^n\times\Omega$为$\boldsymbol{X}$关于$(\Omega,\mathscr{F}_2)$(或$\mathscr{F}_2$)的条件密度—概率分布。

③ 称 $F_{X|\mathscr{F}_2}(x_1,x_2,\cdots,x_n|\omega),(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n,\omega\in\Omega$ 为 X 关于$(\Omega,\mathscr{F}_2)$(或 $\mathscr{F}_2$)的条件密度—分布函数。

(4) X 关于 Y 的条件密度—概率,条件密度—概率分布和条件密度—分布函数

设 $Y=(Y_1,Y_2,\cdots,Y_m)$是随机向量,它的值概率空间为$(\mathbf{R}^m,\mathscr{B}^m,P_Y)$。如果在上段中取 $\mathscr{F}_2=\sigma(Y)$,那么从(3.3.34)开始用$(y_1,y_2,\cdots,y_m)$代替 ω 后所有的结论仍然成立。这时(3.3.34)代之为 $p(C|y_1,y_2,\cdots,y_m;Y),C\in\sigma(X),(y_1,y_2,\cdots,y_m)\in\mathbf{R}^m$,并且在不计 $P(\sigma(Y))$的零概率集意义下成立

$$p(C|Y_1(\omega),Y_2(\omega),\cdots,Y_m(\omega);Y)=p(C|\omega;\sigma(Y)) \quad (3.3.38)$$

(3.3.35)和(3.3.37)分别成为

$$p_X(A|y_1,y_2,\cdots,y_m;Y)\stackrel{\text{def}}{=}p(X\in A|y_1,y_2,\cdots,y_m;Y) \quad (3.3.39)$$

$$F_{X|Y}(x_1,x_2,\cdots,x_n|y_1,y_2,\cdots,y_m)\stackrel{\text{def}}{=}p_X\left(\prod_{i=1}^{n}(-\infty,x_i]|y_1,y_2,\cdots,y_m;Y\right) \quad (3.3.40)$$

定义 3.3.13 ① 称 $p(C|y_1,y_2,\cdots,y_m,Y),C\in\sigma(X),(y_1,y_2,\cdots,y_m)\in\mathbf{R}^m$ 为 X 关于 Y 的条件密度—概率。

② 称 $p_X(A|y_1,y_2,\cdots,y_m;Y),A\in\mathscr{B}^n,(y_1,y_2,\cdots,y_m)\in\mathbf{R}^m$ 为 X 关于 Y 的条件密度—概率分布。

③ 称 $F_{X|Y}(x_1,x_2,\cdots,x_n|y_1,y_2,\cdots,y_m),(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$,$(y_1,y_2,\cdots,y_m)\in\mathbf{R}^m$ 为 X 关于 Y 的条件密度—分布函数。

3.3.6 X 关于子试验$(\Omega,\mathscr{F}_2)$的条件数学期望

给定概率空间$(\Omega,\mathscr{F},P)$。假定 X 是其上的随机向量和$(\Omega,\mathscr{F}_2)$是它的子试验。

定义 3.3.14 设 $p(C|\omega;\mathscr{F}_2),(C,\omega)\in\sigma(X)\times\Omega$ 是 X 关于$(\Omega,\mathscr{F}_2)$的条件密度—概率,如果积分

$$E(\boldsymbol{X}|\mathscr{F}_2)(\omega) \stackrel{\text{def}}{=} \int_{\Omega} \boldsymbol{X}(\omega^*) p(\mathrm{d}\omega^* | \omega; \mathscr{F}_2) \quad (3.3.41)$$

存在,则称 $E(\boldsymbol{X}|\mathscr{F}_2)(\omega)$ 为 $\boldsymbol{X}$ 关于子试验 $(\Omega,\mathscr{F}_2)$(或 $\mathscr{F}_2$)的条件数学期望。

容易证明,当 $\mathrm{E}X$ 存在时 $\mathrm{E}(\boldsymbol{X}|\mathscr{F}_2)$ 必定存在。另外,式(3.3.41)在 $\boldsymbol{X}$ 的值概率空间中的对应形式为

$$\mathrm{E}(\boldsymbol{X}|\mathscr{F}_2)(\omega) = \int_{\mathbf{R}^n} \boldsymbol{x} p_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}|\omega;\mathscr{F}_2) = \int_{\mathbf{R}^n} \boldsymbol{x} \mathrm{d}_x F_{\boldsymbol{X}|\mathscr{F}_2}(\boldsymbol{x}|\omega) \quad (3.3.42)$$

定理 3.3.3 条件数学期望 $\mathrm{E}(\boldsymbol{X}|\mathscr{F}_2)$ 具有性质:

(1) 它是 $\mathscr{F}_2$ 可测的随机向量;

(2) 对任意的 $C\in\mathscr{F}_2$ 成立

$$\int_C \mathrm{E}(\boldsymbol{X}|\mathscr{F}_2)(\omega)P(\mathrm{d}\omega) = \int_C \boldsymbol{X}(\omega)P(\mathrm{d}\omega) \quad (3.3.43)$$

反之,满足(1)和(2)的函数 $\mathrm{E}(\boldsymbol{X}|\mathscr{F}_2)$ 存在,并且在不计 $\mathscr{F}_2$ 中零概率集的意义下惟一。

现代概率论中用该定理中条件(1)和(2)定义条件数学期望 $\mathrm{E}(\boldsymbol{X}|\mathscr{F}_2)$。为了显示它是数学期望的推广,再引进条件密度—概率和条件密度—概率分布,以及条件密度—分布函数,然后证明式(3.3.41)与(3.3.42)成立[25]。

3.3.7 X 关于 Y 的条件数学期望

当 $\mathscr{F}_2=\sigma(\boldsymbol{Y})$ 时我们把 $\mathrm{E}(\boldsymbol{X}|\sigma(\boldsymbol{Y}))$ 改记为 $\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})$。即

$$\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})(\omega) = \mathrm{E}(\boldsymbol{X}|\sigma(\boldsymbol{Y}))(\omega) \quad (3.3.44)$$

转向 $\boldsymbol{Y}$ 的值概率空间 $(\mathbf{R}^m,\mathscr{B}^m,P_{\boldsymbol{Y}})$。由于 $\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})(\omega)$ 是 $\sigma(\boldsymbol{Y})$ 可测函数,应用可测变换的知识得出:存在波莱尔函数 $\boldsymbol{\Psi}(\boldsymbol{y})$,$\boldsymbol{y}\in\mathbf{R}^m$ 使得

$$\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})(\omega)=\boldsymbol{\Psi}[\boldsymbol{Y}(\omega)] \quad (P(\sigma(\boldsymbol{Y}))\text{-a.s.}) \quad (3.3.45)$$

按照概率论的习惯，把 $\Psi(y)$ 改记为 $\mathbf{E}(\boldsymbol{X}|\boldsymbol{Y})(\boldsymbol{y}),\boldsymbol{y}\in\mathbf{R}^m$。注意到(3.3.44)后引进：

定义 3.3.15 称 $\mathrm{E}(X|Y)$ 为 X 关于 Y 的条件数学期望。如果它的自变量为 y，则称为柯尔莫哥洛夫形式的；如果它的自变量为 ω，则称为杜布形式的[①]。

3.3.8 注记

条件数学期望是现代概率论中最重要的概念之一。它在随机过程和数理统计的发展中起关键性的作用。究其原因，大致有

(1) 平均值是一个广泛使用，久经考验的一个普及概念，缺点是适用范围狭窄。数学期望和条件数学期望把它推广到一般场合，成为反映随机变量族统计知识的重要工具。

(2) 假定 $(\Omega,\mathscr{F},P)$ 是用(2.7.4)的方式产生的概率空间。那么 $\mathrm{E}X$ 可以理解为 X 在条件 $\mathscr{C}$ 下的平均值。如果用子试验 $(\Omega,\mathscr{F}_2)$ 加强条件 $\mathscr{C}$（公理 V_2），这时 X 的平均值发生变化，成为条件平均值 $\mathrm{E}(X|\mathscr{F}_2)$。因此，条件数学期望 $\mathrm{E}(X|\mathscr{F}_2)$ 既是某些应用问题寻找的答案，又是许多应用问题寻找答案时的重要工具。

(3) 一般化，当我们得到 $(\Omega,\mathscr{F},P)$ 后依次进行子试验 $(\Omega,\mathscr{F}_1),(\Omega,\mathscr{F}_2),\cdots$ 这时 X 的平均值将不断地变化，得到条件数学期望序列

$$\mathrm{E}X,\mathrm{E}(X|\mathscr{F}_1),\mathrm{E}\{\mathrm{E}(X|\mathscr{F}_1)|\mathscr{F}_2\},\cdots$$

研究这样的序列具有重大的理论和实用意义。

(4) 条件数学期望 $\mathrm{E}(X|\mathscr{F}_2)$ 是随机向量，具有许多性质，有关的计算形成一套比较成熟的方法。

(5) 对任意的 $C\in\mathscr{F}$，由(3.3.41)得出

① 有些书籍把柯氏形式的条件数学期望记为 $\mathrm{E}(X|Y=y)$。参看定义 3.3.11 前的脚注。

$$\mathrm{E}(\chi_C|\mathscr{F}_2)(\omega)=p(C|\omega;\mathscr{F}_2)=p(C|\mathscr{F}_2)(\omega) \tag{3.3.46}$$

于是，事件 C 关于 $\mathscr{F}_2$ 的条件密度等于它的示性函数 χ_C 关于 $\mathscr{F}_2$ 的条件数学期望。

3.4 宽随机过程的基本概念

设 J 是任意无限指标集。

定义 3.4.1 给定概率空间 $(\Omega,\mathscr{F},P)$ 和其上随机变量 $X_\alpha(\omega)(\alpha\in J)$。称随机变量族 $\{X_\alpha(\omega)|\alpha\in J\}$ 为宽随机过程①。简记为 $X_J(\omega)$，或 X_J，或 $X(J)$，或 $X_\alpha(\omega),\alpha\in J$。

定义 3.4.2 给定概率空间 $(\Omega,\mathscr{F},P)$ 和 J 维波莱尔试验 $(\mathbf{R}^J,\mathscr{B}^J)$。假定 $X_J(\omega)\overset{\text{def}}{=}\{X_\alpha(\omega)|\alpha\in J\}$ 是定义在 Ω 上取值于 $\mathbf{R}^J$ 的函数，如果对任意的 $B\in\mathscr{B}^J$ 有

$$(X_J\in B)\overset{\text{def}}{=}\{\omega|X_J(\omega)\in B\}\in\mathscr{F} \tag{3.4.1}$$

则称 $X_J(\omega)$ 为 $(\Omega,\mathscr{F},P)$ 上的宽随机过程。

从 J 维波莱尔试验的定义容易推出两个定义等价。令

$$\mathscr{A}_J=\{(X_J\in A)|A\text{ 是(2.5.33)规定的柱事件，且 }n\geqslant1\} \tag{3.4.2}$$

$$\sigma(X_J)=\{(X_J\in B)|B\in\mathscr{B}^J\} \tag{3.4.3}$$

那么 $\sigma(X_J)$ 是 $\mathscr{F}$ 的子事件 σ 域，$\mathscr{A}_J$ 是生成 $\sigma(X_J)$ 的芽事件集。

定义 3.4.3 称 $\sigma(X_J)$ 为 X_J 产生的事件 σ 域，它是宽过程 X_J 涉及的所有事件组成的集合。

3.4.1 表现概率空间和数字特征

定义 3.4.4 称 $(\Omega,\sigma(X_J),P)$ 为宽过程 X_J 的表现概率空间。

① 这个术语是我们杜撰的。强调它是一个特定的族，以便与它的各种各样的子族相区别。

称$(\Omega,\sigma(X_J))$为 X_J 产生的子试验。

定义 3.4.5 如果对任意的$\alpha\in J$,$\mathrm{E}X_\alpha$存在。则称定义在 J 上的数值函数

$$\mathrm{E}X_J \stackrel{\text{def}}{=} \{\mathrm{E}X_\alpha | \alpha\in J\} \tag{3.4.4}$$

为宽过程 X_J 的数学期望,或期望函数。

定义 3.4.6 如果对任意的 $\alpha,\beta\in J$,协方差 $\mathrm{cov}(X_\alpha,X_\beta)$存在,则称定义在 $J\times J$ 上的数值函数

$$\mathrm{cov}(X_J,X_J) \stackrel{\text{def}}{=} \{\mathrm{cov}(X_\alpha,X_\beta) | \alpha,\beta\in J\} \tag{3.4.5}$$

为宽过程 X_J 的协方差,或协方差函数。

期望函数与协方差函数皆可以用$(\Omega,\mathscr{F},P)$上的积分式表示。它们分别是

$$\mathrm{E}X_J = \int_\Omega X_J(\omega)P(\mathrm{d}\omega) \stackrel{\text{def}}{=} \left\{\int_\Omega X_\alpha(\omega)P(\mathrm{d}\omega) \,|\, \alpha \in J\right\} \tag{3.4.6}$$

$$\mathrm{cov}(X_J,X_J) = \left\{\int_\Omega [X_\alpha(\omega) - \mathrm{E}X_\alpha][X_\beta(\omega) - \mathrm{E}X_\beta]P(\mathrm{d}\omega) \,|\, \alpha,\beta \in J\right\} \tag{3.4.7}$$

3.4.2 值概率空间和有穷维分布函数族

设$(\mathbf{R}^J,\mathscr{B}^J)$是定义 2.4.2 中的波莱尔试验。应用(3.4.3)可以在$(\mathbf{R}^J,\mathscr{B}^J)$上引进集函数

$$P_{X(J)}(B) = P(X_J \in B), B \in \mathscr{B}^J \tag{3.4.8}$$

易证 $P_{X(J)}(\mathscr{B}^J)$是概率测度。由定理 2.8.4 得出:存在相容的有穷维分布函数族

$$F=\{F(\alpha_1,x_1;\alpha_2,x_2;\cdots;\alpha_n,x_n) | n\geqslant 1,\alpha_1,\alpha_2,\cdots,\alpha_n\in J\} \tag{3.4.9}$$

使得 F 与 $P_{X(J)}(\mathscr{B}^J)$用关系式(2.8.26)相互惟一确定。

定义 3.4.7 称$(\mathbf{R}^J,\mathscr{B}^J,P_{X(J)})$为宽过程 X_J 的值概率空间,

且称 $P_{X(J)}(\mathscr{B}^J)$ 为 X_J 的概率分布，F 为 X_J 的有穷维分布函数族。

测度论中可测变换的理论保证宽过程 X_J 在$(\boldsymbol{\Omega},\mathscr{F},P)$上的性质和表达式与$(\mathbf{R}^J,\mathscr{B}^J,P_{X(J)})$的对应性质和表达式之间可以互相转换。即是说，值概率空间包含宽过程的全部统计知识。

在值概率空间中期望函数和协方差函数的表达式分别为

$$\mathrm{E}X_J=\int_{\mathbf{R}^J}x_J P_{X(J)}(\mathrm{d}x_J)\stackrel{\text{def}}{=\!=}\left\{\int_{\mathbf{R}^J}x_\alpha P_{X(J)}(\mathrm{d}x_J)\,|\,\alpha\in J\right\}\tag{3.4.10}$$

$$\mathrm{cov}(X_J,X_J)=\left\{\int_{\mathbf{R}^J}(x_\alpha-\mathrm{E}X_\alpha)(x_\beta-\mathrm{E}X_\beta)P_{X(J)}(\mathrm{d}x_J)\,|\,\alpha,\beta\in J\right\}\tag{3.4.11}$$

应用有穷维分布函数族 F，它们分别简化为

$$\mathrm{E}X_J=\left\{\int_{\mathbf{R}}x\mathrm{d}_xF(\alpha,x)\,|\,\alpha\in J\right\}=\left\{\mathrm{E}X_\alpha\,|\,\alpha\in J\right\}\tag{3.4.12}$$

$$\begin{aligned}\mathrm{cov}(X_J,X_J)&=\{\mathrm{cov}(X_\alpha,X_\beta)\,|\,\alpha,\beta\in J\}\\&=\left\{\int_{\mathbf{R}\times\mathbf{R}}(x-\mathrm{E}X_\alpha)(y-\mathrm{E}X_\beta)\mathrm{d}_x\mathrm{d}_yF(\alpha,x;\beta,y)\right\}\end{aligned}\tag{3.4.13}$$

其中 $F(\alpha,x)=P(X_\alpha\leqslant x)$，$F(\alpha,x;\beta,y)=P(X_\alpha\leqslant x,X_\beta\leqslant y)$分别为一维和二维分布函数。

于是，宽过程找到了表达自己全部统计知识的最方便、最简单的工具——有穷维分布函数族和由它产生的数字特征。

有穷维分布函数族和值概率空间在宽过程的研究中具有重大的意义。对随机变量陈述的理由也适用于宽过程。

目前出现的新情况是：无穷维空间$(\mathbf{R}^J,\mathscr{B}^J)$上的测度和积分理论是抽象的。这个事实虽然增添了研究宽过程的困难，但给无穷维空间的研究带来益处，因为宽过程为研究抽象的无穷维空间提

供一类直观的背景材料。

3.4.3 有穷维子随机向量族

定义 3.4.8 设 X_J 为宽过程。任取 J 中 n 个不同的元素，并赋予顺序。假定赋序后是 $\alpha_1,\alpha_2,\cdots,\alpha_n$，那么称$(X_{\alpha_1},X_{\alpha_2},\cdots,X_{\alpha_n})$为宽过程 X_J 的 n 维子向量。为了反映序关系是松散的，$(X_{\alpha_1},X_{\alpha_2},\cdots,X_{\alpha_n})$又记为$(\alpha_1,X_{\alpha_1};\alpha_2,X_{\alpha_2};\cdots;\alpha_n,X_{\alpha_n})$。

定义 3.4.9 用 fV 表示 X_J 的所有有限维子随机向量组成的集合，即

$$fV=\{(\alpha_1,X_{\alpha_1};\alpha_2,X_{\alpha_2};\cdots;\alpha_n,X_{\alpha_n})\mid n\geqslant 1,\alpha_1,\alpha_2,\cdots,\alpha_n\in J\} \tag{3.4.14}$$

称 fV 为宽过程的有穷维子向量族。

用 X_J 的有穷维子向量族产生：

(1) 子向量$(\alpha_1,X_{\alpha_1};\alpha_2,X_{\alpha_2};\cdots;\alpha_n,X_{\alpha_n})$在概率空间$(\Omega,\mathscr{F},P)$中产生一个子试验$(\Omega,\mathscr{F}_{\alpha_1,\alpha_2,\cdots,\alpha_n})$和一个子表现概率空间$(\Omega,\mathscr{F}_{\alpha_1,\alpha_2,\cdots,\alpha_n},P)$。这里 $\mathscr{F}_{\alpha_1,\alpha_2,\cdots,\alpha_n}=\sigma(X_{\alpha_1},X_{\alpha_2},\cdots,X_{\alpha_n})$。

(2) n 维子向量的值概率空间和概率分布分别称为宽过程的 n 维边缘值概率空间和 n 维边缘概率分布。

(3) 称 n 维子向量的分布函数为 X_J 的 n 维边缘分布函数。显然，(3.4.9)是所有边缘分布函数组成的集合。

(4) 数字特征 $\mathrm{E}X_J$ 和 $\mathrm{cov}(X_J,X_J)$分别由一维和二维边缘分布函数确定。

(5) fV 可以产生许多的条件数学期望和各种形式的条件概率，它们反映子随机向量之间的关系。

3.4.4 指标集 J

如果在指标集 J 上赋予序关系（全序或半序），拓扑关系，代数关系，则得到一些特殊类型的宽随机过程。当指标集为一维欧氏

空间 **R** 的子集 T 时称宽过程 X_T，或 $X_t(\omega)$，$t\in T$ 为随机过程①。随机过程是历史最悠久，统计知识最丰富的宽过程类。其他类型的宽过程研究皆起源于它。

为了增强直观性和形象性，通常把随机过程的指标集 T 解释为时间。于是，人们能够把许许多多的子随机试验，边缘分布函数，各种形式的条件概率，条件数学期望等研究对象按照时间的先后次序组织成条理分明，联系明显的"流"。例如，一类特殊的条件数学期望"流"被称为鞅；又如，人们把一类特殊的子随机试验"流"称为事件流或过滤，引入事件流空间或滤过的概率空间[20, 23, 26]。作者把随机过程定义在事件流空间上，为马尔科夫(A. A. Markov)过程和鞅的研究带来许多的方便[20]。

总之，如何把有穷维子随机向量族 fV 产生的"事物"组织起来，并进行透彻的研究是宽过程首先面临的任务。

数学永远是通过有限去认识无限，宽随机过程也是如此。当我们对 fV 有了初步的认识后，接着的任务是引进收敛性概念，发展极限方法。应用极限方法我们能够通过 fV 的统计知识演绎出宽随机过程或它的无限维子随机变量族的统计知识和统计规律。

本书不涉及这项任务，而是用引出这项诱人的任务结束本书。

---

① 在应用中经常使用它的一种推广形式：$\boldsymbol{X}_T$ 或 $\boldsymbol{X}_t(\omega)$，$t\in T$。称之为随机向量过程，这里 $\boldsymbol{X}_t(\omega)$ 是随机向量。

参 考 文 献

1. 柯尔莫哥洛夫 A H 著．概率论基本概念．丁寿田译．上海：商务印书馆，1952
2. de Finetti B. Probability, Induction and Statistics, The art of guession. John Wiley & Sons Ltd., 1972
3. Rényi A. Probability Theory. New York: American Elsevier, 1970
4. Savage L J. The Foundations of Statistics. New York: John Wiley, 1954
5. Pratt J W, Raiffa H, Schlaifer R. The Foundation of Decision Under Uncertainty: An Elementary Exposition. J. Amer. Stat. Assoc., 1964, 59: 353～375
6. de Groot M. Optimal Statistical Decision. New York: McGraw-Hill, 1970
7. Fishburn P C. Subjective Expected Untility: A Review of Normative Theories. Theory & Decision, 1981, 13: 139～199
8. Fishburn P C. The Axioms of Subjective Probability. Statistical Science, 1986, 1: 335～358
9. 希尔伯特 D. 著．几何基础(上册)，第二版．江泽涵，朱鼎勋译．北京：科学出版社，1987
10. 华东师范大学数学系编．概率论与数理统计教程．北京：高等教育出版社，1984
11. 张尧庭．概率概念的发展和争论．见：邓东皋等编．数学与文化．北京：北京大学出版社，1990
12. Mises R. Wahrscheinlichkeitsrechnung. Leipzig and Wien, 1931
13. Keynes J M. A Treatise on Probability. Landon: MacMillan, 1921
14. 格涅坚科 Б. В. 著．概率论教程．丁寿田译．北京：人民教育出版社，1956
15. Sikorski R. Boolean Algebras. Berlin: Springer-Verlag, 1960
16. Halmos P. R. 著．测度论．王建华译．北京：科学出版社，1958
17. 豪斯道夫 F. 著．集论．张义良，颜家驹译．北京：科学出版社，1960

18. Hanisch H, Hirsch W M, Rényi A. Measures in Denumerable Space. Amer. Math. Monthly, 1969, 76:494～502
19. Gabriel J P, Milasevic P. A Remark on Probability Spaces. J. of Theoretical Probability, 1991, 4:191～195
20. 熊大国．随机过程理论与应用．北京：国防工业出版社，1991
21. 费勒 W. 著．概率论及其应用（上册）．胡迪鹤，林向清译．北京：科学出版社，1964
22. Loève M. Probability Theory. 4th ed. Springer-Verlag, Ⅰ, 1977; Ⅱ, 1978
23. Дынкин Е Б. Марковские Процессы. Фчзматгиз, М., 1963
24. Скороход А В. σ-Алгебри событяй на веродтностных пространствах подобие и факторизация. Теория Вероятн. и её Примен. 1991, 36:127～137
25. 程士宏．高等概率论．北京：北京大学出版社，1996
26. 杜布 J L 著．经典位势论与概率位势论（下册）．杨振明，张润楚等译．北京：科学出版社，1993
27. 王爵渊，王祥麟．遗传学漫谈．南京：江苏科学技术出版社，1982，第16页

[General Information]
书名=概率论自然公理系统-随机世界的数学模型
作者=熊大国
页数=184
SS号=10443356
DX号=
出版日期=2000年11月第1版
出版社=清华大学出版社